

STATISTICAL ANALYSIS OF STOCHASTIC VOLATILITY MODELS

Ph.D. Thesis
Eötvös Loránd University, Faculty of Sciences

ZSANETT ORLOVITS

SUPERVISOR: László Gerencsér, D.Sc.
Scientific Adviser

DOCTORAL SCHOOL OF MATHEMATICS
Director: Miklós Laczkovich, member of HAS
DOCTORAL PROGRAMME OF APPLIED MATHEMATICS
Director: György Michaletzky, D.Sc.



Computer and Automation Institute of the Hungarian Academy of
Sciences, Stochastic Systems Research Group
and
Budapest University of Technology and Economics
Mathematical Institute
Budapest, July 2011

Contents

Table of Contents	2
Acknowledgement	3
1 Introduction	5
1.1 Motivation and our scope	5
1.2 Presentation overview	7
1.3 Notations	10
2 Financial time series: facts and models	11
2.1 Some stylized facts of asset returns	11
2.2 The model class	14
3 Basic properties of GARCH models	20
3.1 State space representation	20
3.2 Stationarity properties	22
3.3 Existence of higher order moments	29
4 Estimation of the parameters of GARCH processes	37
4.1 QMLE of GARCH models	37
4.2 On-line estimation	41
5 The general recursive estimation scheme	46
5.1 The BMP scheme	46
5.2 Basic assumptions	48
5.3 Resetting	54

5.4	Convergence	56
6	Stability properties of block-triangular stationary random matrices	59
6.1	Introduction and the main results	59
6.2	Proofs	65
6.2.1	Proof of Theorem 6.1.1	65
6.2.2	Discussion on the proof of Theorem 6.1.1	72
6.2.3	Proof of Theorem 6.1.2	74
7	Analysis of the recursive algorithm	81
7.1	The algorithm and the main result	81
7.2	Verification of conditions for convergence	85
7.3	Simulation results	89
8	A strong approximation result for GARCH processes	94
8.1	The main result	94
8.2	Proof of Theorem 8.1.2	96
9	Conclusions and further developments	105
A	Basic properties of Kronecker products	108
B	Proof of Lemma 7.2.4 and 7.2.5	110
	References	113
	Summary	122
	Összefoglaló	123

Acknowledgement

I am very glad that there is a page of acknowledgements, for which, in contrast to the rest of this thesis, words come easy.

In the first place I would like to thank my Supervisor and mentor Dr. László Gerencsér. I am especially grateful for his support during the past years, his patience and his belief and trust in me.

My special thanks goes to Prof. György Michaletzky whose excellent teaching provided me a first contact with stochastic processes, Markov chains and system theory.

I would like to express my gratitude to the Hungarian Academy of Sciences (MTA), for the three years scholarship which supported my Ph.D. research I present in this thesis. I am gratefully acknowledge the Computer and Automation Institute of the Hungarian Academy of Sciences (MTA SZTAKI), for providing me with the conditions for my work. I am deeply indebted to the members of the Stochastic Systems Research Group at SZTAKI, especially to Miklós Rásonyi, for the friendly atmosphere and fruitful discussions. My special thanks goes to Prof. Lajos Rónyai for his helpful discussions.

Support from the National Research Foundation of Hungary (OTKA) under Grant no. T047193 is gratefully acknowledged.

I would like to thank the Budapest University of Technology and Economics (BME) for the supporting environment provided me for my research during the last years. Thanks are due to my colleagues at the Mathematical Institute of BME, especially at the Department of Differential Equations, for their help and their backing.

Finally, I would like to thank my family for their patience and love. Without them I could not have written this thesis. Last but not at least I thank all my friends for their help and tolerance. I hope that they are a bit proud that they indirectly also contributed to this thesis.

Chapter 1

Introduction

1.1 Motivation and our scope

Over the last few decades financial mathematics has become one of the few success stories which has attracted the attention of mathematicians, economists, econometricians, physicists, psychologists and many more. The main reason of this success is that the advantage of knowing about risks is that we can change our behavior to avoid them.

This simple concept has a long history in the economics and financial literature, see e.g. the portfolio theory of Markowitz [66], the Capital Asset Pricing Model (CAPM) of Sharp [82] or the option pricing model developed by Black and Scholes [12] and Merton [67]. Parallel to the pricing theory of derivatives, econometricians have modelled financial time series. Financial time series analysis deals with the analysis of data collected on financial markets. The aim of this analysis is to understand and explain the mechanism of highly volatile real-life financial time series such as the log-returns of share prices, foreign exchange rates and stock indices, and to get an acceptable model, which is mathematically tractable. It has been observed by econometricians that a reasonable model for financial processes needs estimates of the variances, its square root called volatility. They immediately recognized that the observed financial time series exhibit a complicated dependence structure, therefore sophisticated mathematical models are needed which should capture many of the so-called stylised facts of financial data such as that the volatilities are changing over time, have heavy

tails and clustering in time. This means that a theory of dynamic volatilities is needed, a fact not true for linear processes.

A first attempt to overcome this problem was the classical paper of Engle [31] where the so-called ARCH (autoregressive conditional heteroscedasticity) model was introduced. The proposed application in that article focused on macroeconomic data. Nowadays, the most widely accepted volatility model in financial application is the standard generalization of ARCH model, called GARCH (generalized ARCH) model, introduced by Bollerslev [14]. The interest of the academic world in these models can be explained by the fact that these models are simple enough for applications, but also rich in theoretical problems. Note that GARCH is just one of the possible ways of nonlinear modelling of financial data. Potential alternative models can be obtained by using e.g. bilinear stochastic systems, see Terdik [88].

The key problem of the statistical analysis of GARCH model is the estimation of the parameters. The literature on estimating the parameters of GARCH models is almost exclusively devoted to off-line quasi-maximum likelihood methods. While the off-line estimation of GARCH models have been analyzed under a variety of conditions in the literature, the on-line or recursive estimation of these processes has attracted little attention until recently.

The main objective of this thesis is to introduce and analyse an adequate on-line or recursive estimation method for the parameters of the standard GARCH models with restricted stability margin under reasonable technical conditions. The main tool in the convergence analysis is an appropriate modification of the theory of recursive estimation within a Markovian framework developed in Benveniste, Metivier and Priouret [7]. In the following we will shortly call this method as BMP-scheme. The idea behind the study of the BMP-scheme is that it reduces the study of individual algorithms to the verification of standard conditions on Markov processes.

The successful adaptation of the BMP-theory has led to another powerful result: using the results of the BMP-scheme and the techniques of Gerencsér [40] we prove a strong approximation theorem for the error term of the off-line maximum likelihood estimator.

The new results of this thesis are based on articles [44, 45, 46]. The first pa-

per is a joint work with György Michaletzky and my supervisor, László Gerencsér, while the second and third paper are joint work with my supervisor, László Gerencsér.

The detailed description of the new results and the structure of the thesis is given in the next section.

1.2 Presentation overview

The thesis consists of the following parts. The introductory section (Chapter 1) contains our motivation, a short summary on the history of modelling financial time series, a detailed overview of the present thesis, and finally the basic notations used throughout the thesis.

The basic properties of general financial time series are considered in Chapter 2. We collect some special stylized facts of financial time series, such as volatility clustering, which are crucial in the model building procedure of financial data. Focusing on these special features we introduce the most widely used class of models in financial applications, namely the ARCH model and its generalization called GARCH model. At the end of this chapter we present some further generalizations of the ARCH and GARCH models which are actually used in the financial literature.

Chapter 3 is devoted to the basic statistical properties of GARCH models. From a technical point of view the key step of this chapter is the state-space representation of GARCH models. This representation transforms the GARCH process into a *linear stochastic system*, which ensures Markovian dynamics. Using this linear dynamics a necessary and sufficient condition for the existence of a strictly stationary solution of GARCH models is given. In the last section of this chapter we address the problem of existence of higher order moments to the solution of GARCH models. As in the case of the existence of a strictly stationary solution to GARCH models, we exploit the connection between GARCH processes and the linear stochastic system given by the state space representation of our process. The stationarity question and the existence of higher order moments of linear stochastic systems are discussed separately, because these properties can be formulated in a more general setting than it is needed for the

case of GARCH processes. A new result related to the existence of higher order moments of a general linear stochastic system is presented. This result is based on the paper of Gerencsér and Orlovits [45].

The objective of Chapter 4 is to discuss the (Gaussian) quasi maximum likelihood estimator for the parameters of the GARCH model, first in an off-line manner. In the second part of the chapter we propose a recursive or on-line method for estimating the parameters of GARCH models and lay the foundations for the analysis of the resulting algorithm. The key point of this section is that the asymptotic estimation problem can be formulated in terms of a linear dynamic with *block-triangular* state matrix.

The construction and the analysis of the proposed recursive algorithm for estimating the GARCH parameters are based on the theory of stochastic approximation with Markovian dynamics presented in Benveniste et al. [7] (BMP-theory), appropriately modified, in particular by applying a suitable resetting mechanism. The basic elements of this general theory will be summarized in Chapter 5.

In order to apply the general BMP-theory for GARCH models we need to introduce some preliminary results on the stability properties of block-triangular stationary random matrix products. The first objective of Chapter 6 is to examine the top-Lyapunov exponent associated with the sequence of block-triangular state matrices. The second purpose of this chapter is to extend the above result to the problem of L_q -stability for the product of independent and identically distributed (i.i.d.) sequences of block-triangular random matrices. This chapter relies on the results presented in Gerencsér, Michaletzky and Orlovits [44] and Gerencsér and Orlovits [45].

The main contribution of Chapter 7 is a rigorous convergence analysis of the recursive estimation method for the parameters of GARCH processes, proposed in Chapter 4, with large stability margin, under reasonable technical conditions. The major achievement of this chapter is a successful adaptation of the BMP-theory to the case of GARCH models. All of the results of this chapter are based on the article of Gerencsér and Orlovits [46]. The viability of the method will be demonstrated by experimental results both for simulated and real data.

In Chapter 8 we investigate a characterization theorem for the error term of

the off-line maximum likelihood estimation using the techniques of Gerencsér [40]. The proof is partially based on the results of the BMP-theory and its application for GARCH processes presented in Chapter 7.

Finally, the concluding chapter gives a summary of the research and outlines directions for further research.

1.3 Notations

The notations used in the thesis are summarized in the following Table.

$\mathbb{N}, \mathbb{Z}, \mathbb{R}$	positive integers, integers, real numbers
$\mathbb{R}^d, \mathbb{R}^{d \times d}$	set of d -dimensional vectors, set of $d \times d$ matrices
$\text{int}M$	interior of the set $M \subset \mathbb{R}^d$
$\log^+ x$	the positive part of $\log x$
C^1	set of functions with continuous derivative
χ_A	indicator function of a set A
EX ($E_Q X$)	the expectation of X (with respect to the measure Q)
$\text{var}(\xi)$	variance of a random variable ξ
$\Pi_\theta(x, A)$	Markov transition kernel of a parametric Markov chain
$\rightarrow, \rightarrow^d$	convergence, convergence in distribution
$N(\mu, \sigma^2)$	normal distribution with expectation μ and variance σ^2
$\mathcal{N}(m, \Sigma)$	multivariate normal distribution with expectation vector m and covariance-matrix Σ
z^{-1}	backward shift operator
$\rho(A)$	spectral radius of a square matrix A
$ x $	any vector norm in \mathbb{R}^d
$\ A\ $	the operator norm of a matrix $A \in \mathbb{R}^{d \times d}$
$\text{vec}A$	operator which takes the columns of a matrix and stacks them column for column in a vector
$A \otimes B$	the Kronecker product of the matrices A and B

Chapter 2

Financial time series: facts and models

In this chapter we collect the most important time series properties of general financial data. These special "stylized facts" are crucial in the model building procedure of financial time series and help us choose an appropriate class of time series processes that can be model financial data. This model class will be introduced in Section 2.2.

2.1 Some stylized facts of asset returns

Consider now a raw financial data which consists of a time series of prices P_n , $n = 1, \dots, N$, of a certain asset (e.g. a stock of a company, a stock index or a foreign currency). Assume that the times of observations are equidistant in order to avoid the difficulties of high-frequency data. Since the price process P_n are believed to be non-stationary, it is a common technique (see Taylor [86]) to transform the observations to so-called *log-returns* by taking log-differences as

$$y_n = \log P_n - \log P_{n-1} = \log \left(1 + \frac{P_n - P_{n-1}}{P_{n-1}} \right).$$

A Taylor-series argument shows that log-returns are close to the relative returns, that is

$$y_n \approx \frac{P_n - P_{n-1}}{P_{n-1}},$$

which describes the relative change over time of the price process.

Figure 2.1 shows a typical display of daily closing prices and log-returns of the Standard and Poor's 500 Composite Stock Price Index (S&P 500) over the period January 1, 1950 through January 28, 2011.¹

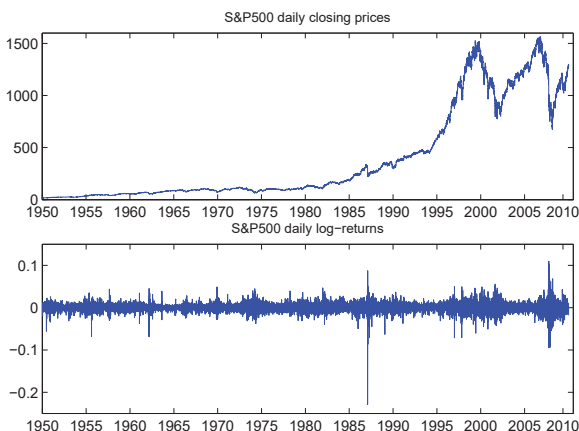


Figure 2.1: S&P 500 daily closing prices and log-returns from January 1, 1950 to January 28, 2011

The study of statistical properties of financial time series has shown a wealth of interesting stylized facts which seem to be common in a wide range of financial applications (see e.g. Cont [23] and Pagan [76]). Here we mention only a few of them: the unconditional distribution of returns display a *heavy tail* with positive excess kurtosis, the *volatility clustering* phenomena, which means that long periods of low volatility are followed by short periods of high volatility, and

¹Information about the composition of this index and historical data were founded by the address <http://www.standardandpoors.com/indices>.

the fact that *autocorrelations of returns are often insignificant*. Researchers have documented these and many other stylized facts about financial time series. For a complete account of these facts we refer the reader to Bollerslev et al. [15].

Among these properties, the phenomenon of volatility clustering has considered in a major way the development of stochastic models in finance. This feature is immediately apparent by looking at Figure 2.1. It can be observed that the volatility clustering feature is seen graphically from the presence of sustained periods of high and low volatilities. Understanding the presence of this phenomenon in financial time series one can observe that, at a basic level, financial price volatility is due to the arrival of new information. That is, volatility clustering is simply clustering of information arrivals, which corresponds to the simple statement that news are clustered in time. On the other hand, there are also several other economic explanations of the presence of this feature, e.g. the heterogeneity of the agent's time scale or the behavioral switching of market participants between fundamentalist and chartists behavior. Further details on these facts can be found e.g. in Cont [24].

The clustering of volatility can be concisely shown by looking at autocorrelations, where volatility clustering will show up as significant autocorrelations in squared or absolute returns. The next figure shows the sample autocorrelation function of log-returns and absolute log-returns for the S&P 500 data set.

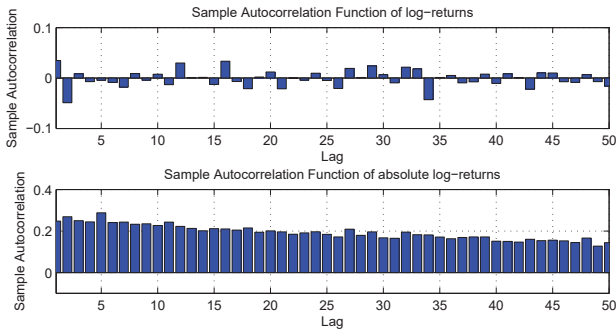


Figure 2.2: Autocorrelations of S&P 500 log-returns and absolute log-returns

2.2 The model class

Modelling financial time series is a complex problem. This complexity is mainly due to the existence of the stylized facts detailed above, which are common to a large number of financial data and are difficult to reproduce by stochastic models.

To look at the classical time series analysis its basic goal is the modelling of the second order structure of the underlying process. In the 1970s the modelling with autoregressive moving average (ARMA) processes became very popular. In the classical time series analysis this ARMA model is important for several reasons: it is mathematically tractable, the determination of the parameter regions of stationarity, causality and invertibility is relatively simple, the autocovariances can be computed explicitly and the parameter estimation is well understood. For a widely used reference on ARMA models see Box and Jenkins [18].

The main reason why ARMA model is not suitable for the description of return data is that the volatility process of ARMA models is constant in time, while the second order properties of financial returns indicates the use of a dynamic volatility structure in order to explain the dependence in the data.

Thus we are forced to model return data by a non-linear process, the minimum requirement for which is to ensure that the conditional variance of the observation process is time-varying. This phenomenon is called *conditional heteroscedasticity*. The first widely accepted non-linear stochastic volatility model is the ARCH model developed by Engle [31]. Here volatility is modelled as the output of a linear finite impulse response (FIR) system, combined with static non-linearities, driven by observed log-returns. In turn, log-returns are assumed to be defined as an i.i.d. process multiplied by the current volatility. Thus we get a stochastic non-linear feedback system, driven by an i.i.d. process. More precisely, (y_n) , with $-\infty < n < +\infty$, is called an ARCH process of order r (ARCH(r)) if it satisfies the equation

$$y_n = \sigma_n \varepsilon_n, \quad (2.1)$$

where σ_n^2 is the conditional variance of y_n given its own past up to time $(n-1)$,

and (ε_n) is an i.i.d. sequence of random variables with zero mean and unit variances, and $\text{E}\sigma_n^2 = \text{E}y_n^2 < \infty$.

A key ingredient of ARCH models is a feedback mechanism in which σ_n^2 is defined in terms of past values y_{n-i}^2 via the linear dynamics

$$(\sigma_n^2 - \gamma^*) = \sum_{i=1}^r \alpha_i^* (y_{n-i}^2 - \gamma^*), \quad n \in \mathbb{Z}, \quad (2.2)$$

with $\gamma^* = \text{E}y_{n-i}^2 = \text{E}\sigma_{n-i}^2 > 0$ and $\alpha_i^* \geq 0$, $i = 1, \dots, r$ denote the true, unknown parameters of the model.

Remark 2.2.1 Note, that the original definition of Engle [31] has been used an alternative parametrization of the model which can be obtained by defining

$$\alpha_0^* = \gamma^* \left(1 - \sum_{i=1}^r \alpha_i^* \right).$$

The parameter restrictions in (2.2) form a necessary and sufficient condition for the positivity of the conditional variance process. It can be easily seen through (2.2) that the ARCH model is a weighted averages of past squared forecast errors, that is, this is a type of weighted variance. These weights could give more influence to recent information and less to the distant past. The big advance of the model is that the weights can be estimated from historical data even though the volatility was never observed. This important discovery on modeling financial data by Engle's ARCH model was recognized by the Nobel Prize in Economics of 2003. The main advance of this model class is that it captures the stylized facts described above relatively well and it is also simple and stationary so that statistical inference is possible.

However many financial applications show that ARCH(r) processes do not fit log-returns very well unless one chooses the order of r quite large. Thus various researchers have thought about improvements. Because the ARCH feedback equation (2.2) bears some resemblance with an autoregressive (AR) structure, the similarity with the ARMA model suggests to introduce an ARMA structure for squared returns. This construction leads us to the so-called generalized ARCH model of order (r, s) (GARCH(r, s)) which was independently introduced

by Bollerslev [14] and Taylor [86] in 1986. The model is written as the multiplicative model (2.1) with specification for the squared conditional variance process σ_n^2 as

$$(\sigma_n^2 - \gamma^*) = \sum_{i=1}^r \alpha_i^* (y_{n-i}^2 - \gamma^*) + \sum_{j=1}^s \beta_j^* (\sigma_{n-j}^2 - \gamma^*), \quad n \in \mathbb{Z}, \quad (2.3)$$

where $\gamma^* = \mathbb{E}y_{n-i}^2 = \mathbb{E}\sigma_{n-j}^2 > 0$ and $\alpha_i^*, \beta_j^* \geq 0$, $i = 1, \dots, r$, $j = 1, \dots, s$.

Remark 2.2.2 Note, that the original definition of Bollerslev [14] has been used an alternative standard parametrization of equation (2.3) which can be received by defining

$$\alpha_0^* = \gamma^* \left(1 - \sum_{i=1}^r \alpha_i^* - \sum_{j=1}^s \beta_j^* \right).$$

Note, that the parameter $\sqrt{\alpha_0^*}$, $\alpha_0^* > 0$, can be recognized as a scale parameter of the process (y_n, σ_n) , $n \in \mathbb{Z}$. This means that if (y_n) is a GARCH(r, s) process with parameters $(\alpha_0^*, \dots, \alpha_r^*, \beta_1^*, \dots, \beta_s^*)^T$, then for any $\lambda > 0$ the process $(\sqrt{\lambda}y_n)$ is a GARCH(r, s) process with parameters $(\lambda\alpha_0^*, \dots, \alpha_r^*, \beta_1^*, \dots, \beta_s^*)^T$ and identical innovations.

An intuitively appealing interpretation of the simplest GARCH(1,1) model is described by Engle [32]: in this model the GARCH forecast variance is a weighted average of three different variance forecast. One is a constant variance which corresponds to the long run average, the second is the forecast made in the previous period and the third is the new information which was not available at the previous step. The weights $\alpha_0^*, \alpha_1^*, \beta_1^*$ determine how fast the variance changes with new information and how fast it reverts to its long-run mean.

The popularity of GARCH models can be explained by some rational arguments. First, it is suggested by construction that the theory behind it might be closely related to the ARMA theory which is widely known, but it is incorrect because of the presence of the squared process (y_n^2) in (2.2). Secondly, using these models one often gets a reasonable fit to real-life financial data. Third, which is maybe the most powerful argument for GARCH, is the fact that its estimation does not provide too many difficulties.

A typical display of a generated GARCH(1, 1) process is seen on Figure 2.2.

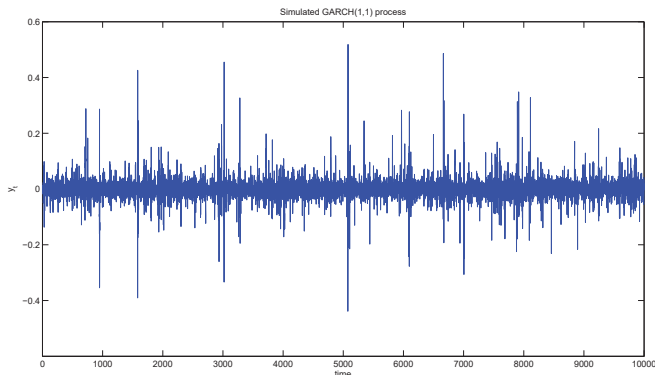


Figure 2.3: Simulated GARCH(1,1) process for 10000 observations with parameters $\alpha_0 = 0.0002$, $\alpha_1 = 0.955$, $\beta_1 = 0.0023$

In many cases of the applications, the basic GARCH model under the normality assumption on the innovation process provides a reasonably good model for analysing the given time series and estimating the conditional volatility process. However, in some cases there are aspects of the model which can be improved so that it can better capture the characteristics and dynamics of a particular time series. For example, if we look at the prices and the corresponding volatilities at the same times on Figure 2.1 it can be observed that the volatility is higher when prices are falling and we can see low volatility period according to a slow and steady growth of prices. This experience on the financial market implies that bad news on the market, i.e. negative shocks, tends to have a larger impact on volatility than good news, i.e. positive shocks. This asymmetric news impact on volatility is often called as the leverage effect of the model, which was first noted by Black [11]. Further details on this observation are given e.g. in Engle [32].

This leverage effect can be incorporated into the GARCH model in several ways. Nelson [73] proposed the so-called *exponential GARCH* (EGARCH) model in the form of (2.1) with the feedback equation

$$\log \sigma_n^2 = \alpha_0 + \sum_{i=1}^r \alpha_i (|\varepsilon_{n-i}| + \psi_i \varepsilon_{n-i}) + \sum_{j=1}^s \beta_j \log \sigma_{n-j}^2, \quad n \in \mathbb{Z}$$

to allow for leverage effects. It can be seen from this equation that no parameter restriction is necessary to ensure the positivity of σ_n^2 , and bad news can have larger impact on volatility by the negativity of the parameter of leverage effects ψ_i .

Another GARCH variant that is capable of modeling leverage effects is the *threshold GARCH* (TGARCH) model proposed by Glosten et al. [48] and Zakoian [96]. In this case the feedback equation has the form

$$\sigma_n^2 = \alpha_0 + \sum_{i=1}^r (\alpha_i y_{n-i}^2 + \psi_i S_{n-i} y_{n-i}^2) + \sum_{j=1}^s \beta_j \sigma_{n-j}^2, \quad n \in \mathbb{Z}$$

where $S_{n-i} = 1$ if $y_{n-i} < 0$ and $S_{n-i} = 0$ if $y_{n-i} \geq 0$. Thus one should expect ψ_i to be positive for bad news to have a larger impact on volatility.

A further generalization of the basic GARCH model which allows the leverage effect is the so-called *power GARCH* (PGARCH(r, d, s)) model developed by Ding et al. [29]. The feedback equation of this model has the form

$$\sigma_n^d = \alpha_1 + \sum_{i=1}^r \alpha_i (|y_{n-i}| + \psi_i y_{n-i})^d + \sum_{j=1}^s \beta_j \sigma_{n-j}^d, \quad n \in \mathbb{Z} \quad (2.4)$$

where d is a positive exponent. The exponent d may also be estimated as an additional parameter of the model which increases the flexibility of the model. As a special case of the PGARCH processes Straumann [84] proposed the so-called *asymmetric GARCH* (AGARCH(r, s)) model which satisfies equation (2.4) with $d = 2$. Ding et al. [29] showed that the PGARCH model also includes many other GARCH variants as special cases.

Further generalizations have been proposed by many researchers. An alphabet soup of these generalized models and the corresponding references can

be found in Engle [32]. See also Franses and Dijk [36] and Terasvirta [87] for excellent surveys of these models.

Chapter 3

Basic properties of GARCH models

In this chapter we summarize the basic statistical properties of the GARCH models. In Section 3.1 the class of these processes and the state space representation of GARCH models are presented briefly. In Section 3.2 we collect the criteria under which weakly and strictly stationary solutions to the GARCH equations exist. The aim of Section 3.3 is to show that under mild conditions GARCH models have finite higher order moments.

3.1 State space representation

Recall from the previous chapter that a time series (y_n) , $n \in \mathbb{Z}$ is called a GARCH(r, s) model if it satisfies the equations

$$y_n = \sigma_n \varepsilon_n, \quad (3.1)$$

$$(\sigma_n^2 - \gamma^*) = \sum_{i=1}^r \alpha_i^* (y_{n-i}^2 - \gamma^*) + \sum_{j=1}^s \beta_j^* (\sigma_{n-j}^2 - \gamma^*), \quad (3.2)$$

where σ_n^2 is the conditional variance of y_n given its own past up to time $(n-1)$, (ε_n) is an i.i.d. sequence of random variables with zero mean and unit variance, $\gamma^* = \mathbb{E}y_{n-i}^2 = \mathbb{E}\sigma_{n-j}^2 > 0$ and $\alpha_i^*, \beta_j^* \geq 0$, $i = 1, \dots, r$, $j = 1, \dots, s$ denote the

true, unknown parameters of the model. Defining the polynomials

$$C^*(z^{-1}) = \sum_{i=1}^r \alpha_i^* z^{-i}, \quad D^*(z^{-1}) = 1 - \sum_{j=1}^s \beta_j^* z^{-j}, \quad (3.3)$$

equation (3.2) can be written in a compact form as

$$D^*(z^{-1})(\sigma^2 - \gamma^*) = C^*(z^{-1})(y^2 - \gamma^*), \quad (3.4)$$

where z^{-1} is the backward shift operator. In the following we will assume that the polynomials C^* and D^* are stable and relative prime.

Let us define the random $(r+s)$ -dimensional state vector X_n^* as

$$X_n^* = (y_n^2, \dots, y_{n-r+1}^2, \sigma_n^2, \dots, \sigma_{n-s+1}^2)^T. \quad (3.5)$$

Then it is easy to see that X_n^* satisfies a linear stochastic system

$$X_{n+1}^* = A_{n+1}^* X_n^* + u_{n+1}^*, \quad n \in \mathbb{Z}, \quad (3.6)$$

with the matrices $A_n^* \in \mathbb{R}^{(r+s) \times (r+s)}$ defined as

$$\left(\begin{array}{cccc|c|cccc|c} \alpha_1^* \varepsilon_n^2 & \alpha_2^* \varepsilon_n^2 & \dots & \alpha_{r-1}^* \varepsilon_n^2 & \alpha_r^* \varepsilon_n^2 & \beta_1^* \varepsilon_n^2 & \beta_2^* \varepsilon_n^2 & \dots & \beta_{s-1}^* \varepsilon_n^2 & \beta_s^* \varepsilon_n^2 \\ 1 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ \hline \alpha_1^* & \alpha_2^* & \dots & \alpha_{r-1}^* & \alpha_r^* & \beta_1^* & \beta_2^* & \dots & \beta_{s-1}^* & \beta_s^* \\ 0 & 0 & \dots & 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 1 & 0 \end{array} \right) \quad (3.7)$$

and $u_n^* \in \mathbb{R}^{r+s}$ defined as

$$u_n^* = (\alpha_0^* \varepsilon_n^2, 0, \dots, 0, \alpha_0^*, 0, \dots, 0)^T.$$

Note, that from the linear dynamics (3.6) it is easy to see that X_n^* is a Markov process, and (A_n^*, u_n^*) , $n \in \mathbb{Z}$ is an i.i.d. sequence of random matrices. This state space representation was introduced by Bougerol and Picard [16] in a slightly different way, however it can be easily seen that the state vector X_n^* and the matrices A_n^*, u_n^* defined here and those of defined by Bougerol and Picard [16] result equivalent state space representation for the GARCH(r, s) process.

3.2 Stationarity properties

The second order properties of the general GARCH model is well-known, see in particular Bollerslev [14]. Recall first that a process (X_n) , $n \in \mathbb{Z}$ is *second-order, or weakly, stationary*, if each X_n is square integrable and if for all $n, m \in \mathbb{Z}$, $E(X_n)$ and $\text{cov}(X_n, X_{n+m})$ are independent of n . The following theorem due to Bollerslev [14] gives a necessary and sufficient condition for the existence of a *second-order stationary* solution of GARCH equations (3.1) and (3.2).

Theorem 3.2.1 (Bollerslev [14]) *The GARCH(r, s) process defined by (3.1) and (3.2) is second-order stationary with*

$$E y_n = 0, \quad \text{Cov}(y_n, y_m) = 0 \quad \text{for } n \neq m$$

and

$$E(y_n^2) = E(\sigma_n^2) = \frac{\alpha_0^*}{1 - \sum_{i=1}^r \alpha_i^* - \sum_{j=1}^s \beta_j^*}$$

if and only if

$$\sum_{i=1}^r \alpha_i^* + \sum_{j=1}^s \beta_j^* < 1. \quad (3.8)$$

Note, that necessity follows trivially by taking expectation in equation (3.2) and noting that $\gamma^* = E y_{n-i}^2 = E \sigma_{n-j}^2 > 0$, $i = 1, \dots, r$, $j = 1, \dots, s$. It is easy to see that a second order stationary solution is necessarily strictly stationary.

Despite the seemingly simple defining equations (3.1) and (3.2), and the easy to handle condition on the existence of a second order stationary solution of GARCH equations, the strictly stationarity conditions of GARCH processes are

not easy to deduce and needs some more sophisticated argumentation. The main idea for tackling the strictly stationarity question of the general GARCH(r, s), model is to write the squared process (y_n^2) in the state space form (3.6) and analyse the resulting linear stochastic system.

Recall first that a process (X_n) , $n \in \mathbb{Z}$ is *strictly stationary* if for all $n, m \in \mathbb{Z}$, the law of $(X_n, X_{n+1}, \dots, X_{n+m})$ is independent of n . Based on the state-space representation of GARCH processes it can be seen that the GARCH equations (3.1) and (3.2) have a unique strictly stationary solution if and only if the linear stochastic system (3.6) has a unique strictly stationary solution with non-negative coordinates. Thus we only have to give an argument for the sufficiency of the latter statement since necessity has been already shown by derivation of (3.6).

The linear stochastic system (3.6) is interesting itself. Several authors have been studied its statistical properties under various set of conditions on the input process. The monograph by Nicholls and Quinn [74] summarizes the basic state of knowledge concerning this model. In the following we summarize the *general theory* of the stationarity question of linear stochastic systems developed by Bougerol and Picard [17].

Let us consider a linear stochastic system given by the state-space equation of the form

$$X_{n+1} = A_{n+1}X_n + u_{n+1}, \quad n \in \mathbb{Z}, \quad (3.9)$$

where $X_n \in \mathbb{R}^d$, A_n is a random matrix in $\mathbb{R}^{d \times d}$ and u_n is a random vector in \mathbb{R}^d . Assume that the following condition holds:

Condition 3.2.1 (A_n, u_n) is a jointly strictly stationary, ergodic sequence of $d \times (d + 1)$ random matrices over some probability space (Ω, \mathcal{F}, P) .

Here Ω is the space or set of elementary events denoted by ω , the σ -algebra \mathcal{F} is the set of measurable subsets of Ω , and P is a probability measure on \mathcal{F} .

A strictly stationary solution (X_n) is called casual if X_{n+1} is measurable with respect to the σ -field $\mathcal{F}_n = \sigma\{A_i, u_i, i \leq n\}$. Both necessary and sufficient conditions for the existence of a strictly stationary casual solution of (3.9) have been given in [17]. To formulate a sufficient condition we need the concept of a

Lyapunov-exponent. Let $|\cdot|$ be any vector norm in \mathbb{R}^d and define an operator norm on the set of $d \times d$ real matrices by

$$\|M\| := \sup_{x \in \mathbb{R}^d, x \neq 0} \frac{|Mx|}{|x|}$$

for $M \in \mathbb{R}^{d \times d}$. Let $\mathcal{A} = (A_n)$ be as above such that

$$\mathbb{E} \log^+ \|A_n\| < +\infty, \quad (3.10)$$

where $\log^+ x$ denotes the positive part of $\log x$. Then it is easy to see that

$$\lambda = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \log \|A_n \dots A_1\| \quad (3.11)$$

exists, where $-\infty \leq \lambda < +\infty$. The proof is based on the observation that $\mathbb{E} \log \|A_n \dots A_1\|$ is subadditive. It also follows that

$$\lambda = \inf_{n > 0} \frac{1}{n} \mathbb{E} \log \|A_n \dots A_1\|. \quad (3.12)$$

The number λ is called the *top-Lyapunov exponent* of \mathcal{A} , and is denoted by $\lambda(\mathcal{A})$. If $A_n = A$ for all n then $\lambda(\mathcal{A})$ is simply the spectral radius of A . A major result of the theory of random matrices is the theorem of Fürstenberg and Kesten stating that

$$\lambda(\mathcal{A}) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \|A_n \dots A_1\| \quad (3.13)$$

almost surely, see Fürstenberg and Kesten [37].

Assume now that (A_n, u_n) satisfies Condition 3.2.1, (3.10) holds for (A_n) and

$$\mathbb{E} \log^+ |u_n| < +\infty. \quad (3.14)$$

Then it is not difficult to show, see Bougerol and Picard [17], that $\lambda(\mathcal{A}) < 0$ implies that (3.9) has a unique strictly stationary and causal solution given by

$$X_n^* = u_n + \sum_{k=1}^{\infty} A_n A_{n-1} \dots A_{n-k+1} u_{n-k}. \quad (3.15)$$

Furthermore, solving (3.9) with any initial condition X_0 forward in time we get that for any $\varepsilon > 0$ we have

$$X_n - X_n^* = O(e^{(\lambda+\varepsilon)n})$$

with probability 1 (w.p.1).

A remarkable *necessary* condition for the existence of a strictly stationary causal solution of (3.9) has been given in the following deep result of Bougerol and Picard [17] for the case when (A_n, u_n) is an i.i.d. sequence.

Theorem 3.2.2 (Bougerol and Picard [17]) *Consider the linear stochastic system given by (3.9), where $X_n^* \in \mathbb{R}^d$, (A_n, u_n) is a jointly strictly stationary sequence of random matrices of size $d \times (d+1)$, jointly satisfying condition (3.10) and (3.14). Let us assume that the sequence (A_n, u_n) is controllable in the sense that there is no proper subspace $V \subset \mathbb{R}^d$, such that*

$$A_0 V + u_0 \subset V \quad \text{w.p.1.}$$

Then if (3.9) has a strictly stationary causal solution (X_n^) , then $\lambda(\mathcal{A}) < 0$.*

The case when the A_n -s have only non-negative entries has been covered in Bougerol and Picard [16]. The proof is far from simple even for the latter case.

For matrices, it may intractable to obtain explicit expressions for the top-Lyapunov exponent and hence to check whether it is strictly negative or not. The following remarks give some observations to obtain tractable sufficient conditions for the negativity of the top-Lyapunov exponent.

Remark 3.2.1 A possible route to establish $\lambda(\mathcal{A}) < 0$ would be use to establish the existence of a causal stationary solution directly, and then use the result of Bougerol and Picard [17] formalized in Theorem 3.2.2 above. Unfortunately,

the direct verification of the existence of a causal stationary solution does not seem to be easy.

Remark 3.2.2 Another way to verify that $\lambda(\mathcal{A}) < 0$ we may use the following observation: if for some $m \geq 1$ we have

$$\mathbb{E}\|A_m \dots A_1\| < 1, \quad (3.16)$$

then $\lambda(\mathcal{A}) < 0$. This follows from the definition of the Lyapunov-exponent given in (3.12), and Jensen's inequality.

Remark 3.2.3 A more delicate condition for $\lambda(\mathcal{A}) < 0$ can be given in the case when the A_n -s are i.i.d. Assume that $\mathbb{E}\|A_1\|^l < +\infty$ for some $l > 0$. Then it is easy to see that for all $0 \leq k \leq l$ the so-called *k-th mean Lyapunov exponent*, see Arnold et al. [3],

$$\lambda_k = \lambda_k(\mathcal{A}) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}\|A_n \dots A_1\|^k \quad (3.17)$$

exists, and $\lambda_k(\mathcal{A}) < +\infty$. Indeed, setting

$$\gamma_n = \log \mathbb{E}\|A_n \dots A_1\|^k,$$

it is directly seen that the sequence γ_n is subadditive, and thus the limit above exists. It is obvious that $\lambda_0(\mathcal{A}) = 0$, and, using Jensen's inequality, we also have

$$\lambda(\mathcal{A}) \leq \lambda_k(\mathcal{A})/k.$$

Thus $\lambda_k(\mathcal{A}) < 0$ implies $\lambda(\mathcal{A}) < 0$. The converse would also follow, if we could prove

$$\lambda(\mathcal{A}) = \frac{d}{dk} \lambda_k(\mathcal{A})|_{k=0}. \quad (3.18)$$

In fact, to prove the converse, a weaker proposition would do. Namely, noting that

$$\lambda_k(\mathcal{A}) = \inf_n \frac{\gamma_n}{n},$$

it is sufficient to show that $\lambda(\mathcal{A}) < 0$ implies $\gamma_n < 0$. But this follows from the fact that for any fixed positive integer n

$$\lambda^{(n)}(\mathcal{A}) = \frac{d}{dk} \lambda_k^{(n)}(\mathcal{A})|_{k=0},$$

where

$$\lambda^{(n)}(\mathcal{A}) = \frac{1}{n} \mathbb{E} \log \|A_n \dots A_1\| \quad \lambda_k^{(n)}(\mathcal{A}) = \frac{1}{n} \log \mathbb{E} \|A_n \dots A_1\|^k,$$

see Lemma 4.8 of Mikosch and Straumann [70]. Thus we arrive at the following result:

Lemma 3.2.1 *Let A_n be an i.i.d. sequence, and assume that $\mathbb{E} \|A_1\|^l < +\infty$ for some $l > 0$. Then $\lambda_k(\mathcal{A}) < 0$ with some $0 < k \leq l$ implies $\lambda(\mathcal{A}) < 0$. Conversely, $\lambda(\mathcal{A}) < 0$ implies $\lambda_k(\mathcal{A}) < 0$ for all sufficiently small k with $0 < k \leq l$.*

For a thorough discussion on the modified Lyapunov-exponent λ_k for arbitrary $k \geq 1$ we refer the reader to Arnold [2] and Arnold et al. [3]. Note, that in the paper of Fang and Loparo [34] it is proved that (3.18) does hold.

Remark 3.2.4 A simple upper bound for the top-Lyapunov exponent $\lambda(\mathcal{A})$ can be obtained by extending the definition of the spectral radius to sets of matrices in a natural way. This leads us to the definition of the joint spectral radius of a set of matrices \mathcal{M} , introduced by Rota and Strang [81], which measures the maximal asymptotic growth rate of long products of matrices taken from \mathcal{M} . More formally, the joint spectral radius is defined as

$$\rho(\mathcal{M}) := \limsup_{k \rightarrow \infty} \rho_k(\mathcal{M}),$$

where

$$\rho_k(\mathcal{M}) = \sup_{A_1, \dots, A_k \in \mathcal{M}} \|A_k \dots A_1\|^{1/k}.$$

It can be easily seen that if the set of matrices consists only on matrix A , the joint spectral radius coincides with the usual notion of spectral radius of a single

matrix

$$\rho(A) = \lim_{n \rightarrow \infty} \|A^n\|^{1/n}.$$

As for the single case, the joint spectral radius does not depend on the matrix norm used. For the properties and the computation of $\rho(\mathcal{M})$ a sequence of useful results has been given in Barabanov [5], Blondel and Nesterov [13], Tsitsiklis [90] and the references therein. It is then obvious that

$$\lambda(\mathcal{M}) \leq \log \rho(\mathcal{M}).$$

Summarizing the above facts and applying for GARCH processes we get the following theorem due to Bougerol and Picard [16]:

Theorem 3.2.3 (Bougerol and Picard [16]) *When $\alpha_0^* > 0$, the GARCH equations (3.1) and (3.2) has a strictly stationary and ergodic solution if and only if the top-Lyapunov exponent associated with the matrices $(A_n^*), n \in \mathbb{Z}$ defined in (3.7) is strictly negative. This solution is unique, and the random vector X_n^* defined in (3.5) satisfies (3.15).*

It is easy to see that a second order stationary solution is necessarily strictly stationary. However, if the noise sequence has finite variance, the stability condition (3.8) gives a handy sufficient condition for the GARCH process to have a strictly stationary solution, which is easy to check. These are summarized in the following Corollary.

Corrolary 3.2.1 (Bollerslev [14], Bougerol and Picard [16]) *For the GARCH(r, s) process with the driving noise sequence $(\varepsilon_n), n \in \mathbb{Z}$ having zero mean and unit variance the following hold:*

- (a) *If $\alpha_0^* > 0$ and $\sum_{i=1}^r \alpha_i^* + \sum_{j=1}^s \beta_j^* < 1$, then the GARCH(r, s) process admits a unique strictly stationary solution.*
- (b) *When $\alpha_0^* > 0$, if the GARCH(r, s) model has a strictly stationary solution, then $\sum_{j=1}^s \beta_j^* < 1$.*

The first proposition of the Corollary is a simple application of Theorem 3.2.3. Indeed, if the stability condition (3.8) holds, then EA_0^* is sub row-stochastic,

hence, by the Perron-Frobenius theorem the spectral radius of EA_0^* satisfies $\rho(EA_0^*) < 1$, which in turn implies $\lambda(\mathcal{A}) \leq \log \rho(EA_0^*) < 0$, (the latter inequality being not quite trivial, for details see Theorem 2 of Kesten and Spitzer [57]), from which the existence of a strictly stationary and causal solution follows.

3.3 Existence of higher order moments

It is often not only necessary for the model in consideration to have a stationary and ergodic solution, but also that it has finite moments of appropriate order. In this section we address the problem of existence of higher order moments to the solution of GARCH models. As in the case of the existence of a strictly stationary solution to GARCH models, we exploit the connection between GARCH processes and the linear stochastic system. Note that the proposed method will be also appropriate in time series framework whenever the underlying time series model can be given in Markovian structure.

The existence of moments for the linear stochastic system have been studied by Feigin and Tweedie [35] using the Markov chain approach. For further results on these questions under various set of conditions on the input process we refer the reader to Brandt [19], Karlsen [56], Pham [78] and Vervaat [93], as well.

Consider again the linear stochastic system defined as

$$X_{n+1} = A_{n+1}X_n + u_{n+1}, \quad n \in \mathbb{Z}, \quad (3.19)$$

where (A_n) is a sequence of random $d \times d$ matrices and (u_n) is a sequence of random d -vectors. Assume that Condition 3.2.1 holds and suppose furthermore, that

$$\mathbb{E}|u_n|^q < +\infty \quad \text{for some } q \geq 1.$$

We have seen in the previous section that, under condition (3.10), the negativity of the top-Lyapunov exponent $\lambda(\mathcal{A})$ implies that (3.19) has a unique strictly stationary and causal solution given by

$$X_n^* = u_n + \sum_{k=1}^{\infty} A_n A_{n-1} \dots A_{n-k+1} u_{n-k}. \quad (3.20)$$

We can then ask under what conditions will the infinite sum on the right hand side of equation (3.20) converge in L_q . To simplify the discussion we will use the following condition introduced by Pham [78]:

Condition 3.3.1 (A_n) is an i.i.d. sequence of random matrices, and the two σ -fields

$$\mathcal{F}_{n+}^A = \sigma\{A_i : i > n\}$$

$$\mathcal{F}_{n-}^u = \sigma\{u_i : i \leq n\}$$

are independent for any n .

Taking the L_q -norm of both sides of (3.20) and applying the triangle inequality we get

$$\mathbb{E}^{1/q}|X_n^*|^q \leq \mathbb{E}^{1/q}|u_n|^q + \sum_{k=1}^{\infty} \mathbb{E}^{1/q}|A_n \dots A_{n-k+1}u_{n-k}|^q. \quad (3.21)$$

Now $|A_n \dots A_{n-k+1}u_{n-k}| \leq \|A_n \dots A_{n-k+1}\| \cdot |u_{n-k}|$, where $\|\cdot\|$ denotes the operator norm in $\mathbb{R}^{d \times d}$. By Condition 3.3.1, the two terms in the latter product are independent, the same being true for their q th power, thus we have for all k

$$\mathbb{E}|A_n \dots A_{n-k+1}u_{n-k}|^q \leq \mathbb{E}\|A_n \dots A_{n-k+1}\|^q \cdot \mathbb{E}|u_{n-k}|^q. \quad (3.22)$$

Taking into account (3.21) and (3.22) yields

$$\mathbb{E}^{1/q}|X_n^*|^q \leq \mathbb{E}^{1/q}|u_n|^q + \sum_{k=1}^{\infty} \mathbb{E}^{1/q}\|A_n \dots A_{n-k+1}\|^q \cdot \mathbb{E}^{1/q}|u_{n-k}|^q.$$

Thus, to prove convergence in L_q in (3.20) for any L_q bounded u_n it is therefore sufficient to show that

$$\mathbb{E}\|A_n \dots A_{n-k+1}\|^q = \mathbb{E}\|A_k \dots A_1\|^q,$$

is summable. Assuming $\mathbb{E}\|A_1\|^q < \infty$, summability implies $\mathbb{E}\|A_k \dots A_1\|^q < 1$ for some k , which in turn implies $\lambda_q(\mathcal{A}) < 0$. Conversely, $\lambda_q(\mathcal{A}) < 0$ implies the required summability. (For the definition and properties of $\lambda_q(\mathcal{A})$ see Remark 3.2.3.) Thus we get that $\mathbb{E}\|A_k \dots A_1\|^q$ is summable over k if and only if

$\lambda_q(\mathcal{A}) < 0$. This brings us to the following definition, introduced in Gerencsér and Orlovits [45]:

Definition 3.3.1 *Let A_n be an i.i.d. sequence, and assume that $E\|A_1\|^q < +\infty$ for some $q \geq 1$. We say that \mathcal{A} is L_q -stable if $\lambda_q(\mathcal{A}) < 0$.*

The existence of finite even order moments of X_n^* was studied by Feigin and Tweedie [35], using Markov-chain techniques, under the following condition, which is more restrictive than Condition 3.3.1:

Condition 3.3.2 *(A_n, u_n) is an i.i.d. sequence of random matrices and (A_n) and (u_n) are also independent of each other.*

Let q denote some positive even integer. They have shown that, under Condition 3.3.2 a sufficient condition for the convergence of (3.20) in L_q and the existence of the q -th moment of X_n^* is that all the eigenvalues of the matrix $E(A_1^{\otimes q})$ are less than unity in modulus, where $A^{\otimes q}$ denotes the q -th Kronecker power of A . (The basic properties of the Kronecker product are given in Appendix A.) We will provide a simple proof of this result under the weaker Condition 3.3.1, see Proposition 3.3.1 below. We will also see that the condition $\rho[E(A_1^{\otimes q})] < 1$, with $\rho(\cdot)$ denoting the spectral radius, is sufficient for $E^{1/q}\|A_n \dots A_1\|^q$ to converge to zero exponentially fast.

It is not known if the condition $\rho[E(A_1^{\otimes q})] < 1$ is also necessary in the context discussed by Feigin and Tweedie [35]. They conjectured that *necessity* indeed holds. The best result we are aware of is the necessity of the above condition for GARCH-processes, see Ling and McAleer [62], reformulated in Theorem 3.3.3 below. However the authors point out that they heavily exploit the special structure of EA_n , and their method does not extend to the case of general state matrices.

An alternative condition for the L_q -stability of the product $A_n \dots A_1$ has been given by Hasminskii [53]. In fact his result is formulated for any *real*, not necessarily integer $q > 0$. He has shown that, assuming that (A_n) is an i.i.d. sequence, a sufficient condition for

$$E\|A_n \dots A_1\|^q \rightarrow 0 \quad \text{as } n \rightarrow \infty \quad (3.23)$$

to hold is that there exists a positive definite function $f(x)$, which is homogeneous of degree q , such that the function

$$Ef(Ax) - f(x)$$

is negative definite. This condition is also necessary, and in fact the value of $Ef(Ax) - f(x)$ can be prescribed: for any positive definite function $g(x)$, which is homogeneous of degree q the equation $Ef(Ax) - f(x) = -g(x)$ has a unique positive definite solution $f(x)$, which is homogeneous of degree q . If q is an even integer, then f and g can be restricted to positive definite q -forms. For details see Hasminskii [53]. We note in passing that (3.23) implies $\lambda_q(\mathcal{A}) < 0$, hence the rate of convergence in (3.23) is geometric. The relationship between the conditions of Feigin and Tweedie and that of Hasminskii will be briefly discussed in Remark 3.3.1.

Now we will provide a simple proof of the key step in deriving the result of Feigin and Tweedie [35] under Condition 3.3.1, which is weaker than Condition 3.3.2.

Theorem 3.3.1 *Let (A_n) be an i.i.d. sequence of random matrices such that $\|A_1\| \in L_q$. Assume that for some even integer $q \geq 2$*

$$\rho [E(A_1^{\otimes q})] < 1. \tag{3.24}$$

Then

$$\lambda_q = \lambda_q(\mathcal{A}) = \lim_{n \rightarrow \infty} \frac{1}{n} \log E\|A_n \dots A_1\|^q < 0.$$

It follows that for any $\varepsilon > 0$ we have

$$E\|A_n \dots A_1\|^q \leq Ce^{(\lambda_q + \varepsilon)n}$$

with some $C = C(\varepsilon) > 0$.

The above theorem is a direct corollary of the following one, both appeared in Gerencsér and Orlovits [45].

Proposition 3.3.1 *Let (A_n) be an i.i.d. sequence of random matrices as above. Then*

$$\mathbb{E}\|A_n \dots A_1\|^q = (\text{vec} I^T) \cdot [\mathbb{E}(A_1^{\otimes q})]^n \cdot (\text{vec} I),$$

where I is a unit matrix of appropriate dimension.

Remark 3.3.1 It is not clear if the equivalence of the above condition and the necessary and sufficient condition of Hasminskii described in terms of q -forms, can be easily established directly. An exception is the case $q = 2$. In this case Hasminskii's condition reduces to the existence of positive definite matrices F and G such that

$$F = \mathbb{E}[A^T F A] + G,$$

which is directly seen to be equivalent to $\rho[\mathbb{E}(A_1^{\otimes 2})] < 1$.

Proof of Proposition 3.3.1: Let

$$\Delta_n = A_n \dots A_1,$$

and let $q = 2k$. By Lemma A.0.6,(v) (see in Appendix A) we have

$$\|A_n \dots A_1\|^q = \|\Delta_n\|^{2k} = (\|\Delta_n\|^k)^2 = \|\underbrace{\Delta_n \otimes \dots \otimes \Delta_n}_{k \text{ times}}\|^2. \quad (3.25)$$

Now for any matrix C we have $\|C\|^2 = \text{tr}(CC^T)$. Taking $C = \Delta_n \otimes \dots \otimes \Delta_n$, and applying Lemma A.0.6,(ii), (iii) we get that

$$\begin{aligned} \|\Delta_n \otimes \dots \otimes \Delta_n\|^2 &= \text{tr}[(\Delta_n \otimes \dots \otimes \Delta_n) \cdot (\Delta_n \otimes \dots \otimes \Delta_n)^T] = \\ &= \text{tr}(\Delta_n \Delta_n^T \otimes \dots \otimes \Delta_n \Delta_n^T). \end{aligned} \quad (3.26)$$

The dynamics of

$$V_n = \Delta_n \Delta_n^T \otimes \dots \otimes \Delta_n \Delta_n^T$$

is obtained by taking into account that

$$\Delta_n = A_n \Delta_{n-1}.$$

Applying Lemma A.0.6,(ii), (iii) in the opposite direction we get

$$\begin{aligned}\Delta_n \Delta_n^T \otimes \dots \otimes \Delta_n \Delta_n^T &= (A_n \Delta_{n-1} \Delta_{n-1}^T A_n^T) \otimes \dots \otimes (A_n \Delta_{n-1} \Delta_{n-1}^T A_n^T) \\ &= (A_n^{\otimes k}) \cdot (\Delta_{n-1} \Delta_{n-1}^T)^{\otimes k} \cdot (A_n^T)^{\otimes k},\end{aligned}\quad (3.27)$$

in short

$$V_n = (A_n^{\otimes k}) \cdot V_{n-1} \cdot (A_n^T)^{\otimes k}. \quad (3.28)$$

Applying the vec operation and using Lemma A.0.6,(vi) we get that

$$\begin{aligned}\text{vec} V_n &= [(A_n^{\otimes k}) \otimes (A_n^{\otimes k})] \cdot \text{vec} V_{n-1} \\ &= (A_n^{\otimes 2k}) \cdot \text{vec} V_{n-1}.\end{aligned}\quad (3.29)$$

Iterating this equation yields

$$\text{vec} V_n = (A_n^{\otimes 2k}) \cdot (A_{n-1}^{\otimes 2k}) \dots (A_1^{\otimes 2k}) \cdot \text{vec} I, \quad (3.30)$$

where $I = V_0$ denotes the identity matrix of appropriate size. Taking the expectation of both sides, and using the fact that (A_n) is an i.i.d. sequence we get

$$\mathbb{E} \text{vec} V_n = \text{vec} \mathbb{E} V_n = [\mathbb{E}(A_1^{\otimes 2k})]^n \cdot \text{vec} I. \quad (3.31)$$

Using Lemma A.0.6,(vii) we conclude that

$$\text{tr} \mathbb{E} V_n = \text{vec} I^T \cdot [\mathbb{E}(A_1^{\otimes 2k})]^n \cdot \text{vec} I, \quad (3.32)$$

from which the claim follows. ■

Remark 3.3.2 An another condition for the existence of higher order moments of (3.19) can also be formalized by using the symmetric tensor power of A_n . For

this purpose, define $(\mathbb{R}^d)^{\odot s}$ as the space of vectors

$$u = (u_{i_1 \dots i_s}, i_j = 1, \dots, d)$$

which are invariant with respect to any permutation of the subscripts of their components. Clearly, $(\mathbb{R}^d)^{\odot 2}$ can be identified with the space of symmetric matrices of order d . For $x \in \mathbb{R}^d$, the s -th symmetric tensor power $x^{\odot s}$ is the vector of $(\mathbb{R}^d)^{\odot s}$ with components $x_{i_1 \dots i_s}$. If $A = (a_{ij})$ is a square matrix of order d , the s -th symmetric tensor power $A^{\odot s}$ is defined as the operator on $(\mathbb{R}^d)^{\odot s}$ which maps the vector u to the vector $A^{\odot s}u = v$ as

$$v_{i_1 \dots i_s} = \sum_{j_1 \dots j_s} a_{i_1 j_1} \dots a_{i_s j_s} u_{j_1 \dots j_s}.$$

The following theorem due to Pham [78] gives a sufficient condition for the existence of higher order moments of the linear stochastic system (3.19).

Theorem 3.3.2 (Pham [78]) *Let us assume that Condition 3.2.1 and Condition 3.3.1 hold and*

$$\rho[E(A_1^{\odot 2m})] < 1. \quad (3.33)$$

Then $E(\|X_n\|^{2m}) < \infty$.

Remark 3.3.3 Carrasco and Chen [20] use also an another condition to establish the existence of higher order moments of the solution of (3.19). It is obvious that, while Condition 3.3.1 is standard, its verification is not trivial. Therefore Carrasco and Chen replace it by the following alternative condition:

Condition 3.3.3 *There exists some even integer $s \geq 2$ such that*

$$E[(\rho(A_1))^s] < 1 \quad \text{and} \quad E\|u_1\|^s < \infty.$$

They proved the following alternative proposition for the existence of higher order moments of (3.19).

Proposition 3.3.2 (Carrasco and Chen [20]) *Assume that Condition 3.2.1 and Condition 3.3.3 hold and*

$$\rho(EA_1) < 1.$$

Then $E(\|X_n\|^s) < \infty$.

The condition for the finiteness of higher-order moments of the general GARCH model has been studied by several authors, as well. In the case of first order ARCH model Engle [31] obtained conditions for the existence of higher order moments. The generalization of this result for ARCH(q) models has been given by Milhoj [72]. In his seminal paper [14] Bollerslev has given condition for the finiteness of moments for the GARCH(1, 1) model. Stability of various forms of GARCH(1, 1) models are studied by He and Terasvirta [54], Carrasco and Chen [20], Karanasos [55] and Zdrozny [95].

The next result on the existence of higher order moments of GARCH(r, s) models has been given in the paper of Ling and McAleer [62].

Theorem 3.3.3 (Ling and McAleer [62]) *The necessary and sufficient condition for $E(y_n^{2m}) < \infty$ is*

$$\rho[E((A_n^*)^{\otimes m})] < 1. \quad (3.34)$$

The sufficiency comes from Theorem 6.1 of Ling [61] and from Proposition 3.3.1, as well. The proof of necessity is given in Ling and McAleer [62]. Note that proof of Theorem 3.3.3 makes full use of the special structure of the GARCH state matrix, and unfortunately this method cannot be applicable to the general linear model defined by (3.9).

Chapter 4

Estimation of the parameters of GARCH processes

The main contribution of this chapter is to discuss the (Gaussian) quasi maximum likelihood estimator of the parameters of the GARCH(r, s) model, and to construct an on-line or recursive estimation method for the parameters of this model. At the end of this chapter we lay the foundations for the analysis of the recursive algorithm for GARCH processes. The rigorous convergence analysis of the proposed recursive algorithm will be given in Chapter 7.

4.1 QMLE of GARCH models

Gaussian quasi-maximum likelihood estimation, i.e., likelihood estimation under the hypothesis of Gaussian innovations, is a popular method which is widely used for inference in time series models. Often it is however a non-trivial task to establish the consistency and asymptotic normality of the quasi-maximum likelihood estimator applied to specific models, and an in-depth analysis of the probabilistic structure generated by the model is called for. A classical example of this kind is the seminal paper by Hannan [52] on estimation in linear ARMA time series.

The estimation, or identification, of the parameters of GARCH processes has attracted considerable attention recently, see e.g. Berkes et al. [10], Lee and Hansen [60], Lumsdaine [65], Mikosch and Straumann [71] and Wiess [94].

All the cited works consider the so-called off-line estimation problem, when the collection of data and statistical analysis are separated in time. The weakest conditions for the strong consistency of an off-line quasi-maximum likelihood method has been given in Berkes et al. [10].

In what follows, we show how one can construct the quasi maximum likelihood estimator of GARCH(r, s) processes. For the identification of the parameters of GARCH processes we proceed similarly to ARMA processes. Write $\theta = (\alpha_0, \alpha_1, \dots, \alpha_r, \beta_1, \dots, \beta_s)^T$ and let $K \subset \mathbb{R}^{r+s+1}$ denote the set of θ -s such that the stability condition

$$\sum_{i=1}^r \alpha_i + \sum_{j=1}^s \beta_j < 1$$

holds and the polynomials C and D , defined by these θ -s, are stable. Let $K_0 \subset \text{int } K$ be a compact domain such that $\theta^* \in \text{int } K_0$. For a fixed tentative value of the system parameters, say θ , we *invert* the system

$$\bar{\sigma}_n^2(\theta) - \gamma = \sum_{i=1}^r \alpha_i (y_{n-i}^2 - \gamma) + \sum_{j=1}^s \beta_j (\bar{\sigma}_{n-j}^2(\theta) - \gamma) \quad (4.1)$$

to get the *frozen parameter* process $\bar{\sigma}_n^2(\theta)$, using the initial values

$$y_n = 0 \quad \text{and} \quad \bar{\sigma}_n^2(\theta) - \gamma = 0 \quad \text{for all } n \leq 0.$$

Note that the initial values are asymptotically irrelevant to the estimation procedure, see e.g. Straumann [84]. Then we compute the estimated driving noise $\bar{\varepsilon}_n(\theta)$ by the inverse equation

$$\bar{\varepsilon}_n(\theta) = \frac{y_n}{\bar{\sigma}_n(\theta)} \quad (4.2)$$

with $n \geq 0$. Now, our task is to compute the log-likelihood function

$$\log f(y_1, \dots, y_N; \theta)$$

for a given data set (y_1, \dots, y_N) . To do this let us consider the general decom-

position

$$\begin{aligned} & \log f(y_1, \dots, y_N | y_0, \dots, y_{-r+1}, \sigma_0^2, \dots, \sigma_{-s+1}^2; \theta) = \\ & = \sum_{n=1}^N \log f(y_n | y_{n-1}, \dots, y_{-r+1}, \sigma_0^2, \dots, \sigma_{-s+1}^2; \theta). \end{aligned}$$

Assuming now that ε_n is standard normal, the (conditional) log-likelihood function, modulo a constant, can be expressed via $\bar{\sigma}_n^2(\theta)$ as

$$L_N = \sum_{n=1}^N l_n = \sum_{n=1}^N -\frac{1}{2} \left(\log \bar{\sigma}_n^2(\theta) + \frac{y_n^2}{\bar{\sigma}_n^2(\theta)} \right). \quad (4.3)$$

The right hand side depends on θ^* via (y_n^2) . To stress dependence of L_N on both θ and θ^* , we shall write $L_N = L_N(\theta, \theta^*)$. The same cost function is used also in the non-gaussian case, see e.g. Berkes and Horváth [9]. Then the general abstract estimation problem, assuming for a moment *stationary initialization*, is to find the solution of the non-linear algebraic equation

$$\mathbb{E}_{\theta^*} \frac{\partial}{\partial \theta} l_0(\theta, \theta^*) = 0. \quad (4.4)$$

The conditional quasi-maximum likelihood estimation $\hat{\theta}_N$ of θ^* is defined as the solution of the equation

$$\frac{\partial}{\partial \theta} L_N(\theta, \theta^*) = L_{\theta N}(\theta, \theta^*) = 0. \quad (4.5)$$

The differentiation here is taken in the almost sure sense. More exactly: $\hat{\theta}_N$ is a random vector such that $\hat{\theta}_N \in K$ for all ω , and if the equation (4.5) has a unique solution in K , then $\hat{\theta}_N$ is equal to this solution. By the measurable selection theorem such a random variable exists.

Define the asymptotic cost function, (a negative log-likelihood for the gaussian case) as

$$W(\theta, \theta^*) = \lim_{n \rightarrow \infty} \mathbb{E} \left[\frac{1}{2} \left(\log \bar{\sigma}_n^2(\theta) + \frac{y_n^2}{\bar{\sigma}_n^2(\theta)} \right) \right]. \quad (4.6)$$

Then the the asymptotic estimation problem is

$$\frac{\partial}{\partial \theta} W(\theta, \theta^*) = \lim_{n \rightarrow \infty} E \frac{\bar{\sigma}_{\theta,n}(\theta)}{\bar{\sigma}_n^3(\theta)} \left(1 - \frac{y_n^2}{\bar{\sigma}_n^2(\theta)} \right) = 0. \quad (4.7)$$

The existence of the above limits can be easily proven, see e.g. Berkes et al. [10]. Also note that, by Lemma 5.5 of Berkes et al. [10], θ^* is the a unique solution of the asymptotic problem in K_0 .

We now list the weakest conditions under which consistency and asymptotic normality of the estimator hold. These conditions are formulated by Berkes et al. [10]:

Condition 4.1.1 *Assume that*

- (i) *the system noise process $(\varepsilon_n), n \in \mathbb{N}$ is an i.i.d. process with zero mean and unit variance and the distribution of ε_0 is not concentrated in two points,*
- (ii) *there exists some $\delta > 0$ such that $E(|\varepsilon_n|^2)^{1+\delta} < +\infty$,*
- (iii) *there is $\mu > 0$ such that $P(|\varepsilon_0| \leq t) = o(t^\mu)$ as $t \downarrow 0$.*

We are now ready to quote Theorems 4.1 and 4.2 of Berkes et al. [10]:

Theorem 4.1.1 (Berkes, Horváth and Kokoszka [10]) *Let $(y_n), n \in \mathbb{N}$ be a stationary GARCH(r, s) process with true parameter vector $\theta^* \in \text{int } K_0$. Assume that the polynomials C^* and D^* are stable, and suppose that Condition 4.1.1 holds. Then the QMLE estimator $\hat{\theta}_N$ is strongly consistent, i.e.*

$$\hat{\theta}_N \rightarrow \theta^* \quad \text{a.s.} \quad \text{as } N \rightarrow \infty.$$

If in addition $E|\varepsilon_0|^4 < +\infty$, the QMLE estimator $\hat{\theta}_N$ is also asymptotically normal, i.e.

$$\sqrt{n}(\hat{\theta}_N - \theta^*) \rightarrow^D \mathcal{N}(0, F_0^{-1} G_0 F_0^{-1}) \quad \text{as } N \rightarrow \infty,$$

where the $(r + s + 1) \times (r + s + 1)$ matrices F_0 and G_0 are given by

$$F_0 = E(l_{\theta\theta,0}(\theta^*)) = 2 \cdot E \left(\frac{\sigma_{\theta,0}(\theta^*) \sigma_{\theta,0}(\theta^*)^T}{\sigma_0^2(\theta^*)} \right) \quad (4.8)$$

$$G_0 = \text{cov}(l_{\theta,0}(\theta^*)) = \frac{E(\varepsilon_0^4 - 1)}{4} E \left(\frac{\sigma_{\theta,0}(\theta^*) \sigma_{\theta,0}(\theta^*)^T}{\sigma_0^2(\theta^*)} \right). \quad (4.9)$$

Remark 4.1.1 Note that Berkes et al. [10] even require $E|\varepsilon_0|^{4+\delta} < \infty$ for some $\delta > 0$, which seems to be however too restrictive since their proof goes through under the weaker condition $E\varepsilon_0^4 < \infty$.

Remark 4.1.2 Elaborating the expressions (4.8) and (4.9) of Theorem 4.1.1 and taking into account

$$R^* = \frac{\partial^2}{\partial \theta^2} W(\theta, \theta^*)|_{\theta=\theta^*} = \lim_{n \rightarrow \infty} 2 E \frac{\bar{\sigma}_{\theta,n}(\theta) \bar{\sigma}_{\theta,n}(\theta)^T}{\bar{\sigma}_n^2(\theta)}|_{\theta=\theta^*},$$

we get, with $(\bar{\sigma}_n(\theta))$ assumed to be stationary, that the asymptotic covariance matrix, say $\Sigma(\theta^*)$, of the off-line estimator is given by

$$\Sigma(\theta^*) = \frac{1}{4} \left(E \frac{\bar{\sigma}_{\theta,n}(\theta^*) \bar{\sigma}_{\theta,n}(\theta^*)^T}{\bar{\sigma}_n^2(\theta^*)} \right)^{-1} \cdot E(\varepsilon_n^4 - 1). \quad (4.10)$$

In the Gaussian case we have $E(\varepsilon_n^4 - 1) = 2$, and $\Sigma(\theta^*)$ is the inverse of the Fisher information matrix. The above expression can also be used to compute the asymptotic covariance matrix or the inverse of the Fisher information matrix empirically. This observation is used in the simulations.

4.2 On-line estimation

The literature on the identification of GARCH models is almost exclusively devoted to *off-line* quasi-maximum likelihood methods. However, given that financial time series are often sampled at high frequency, a more convenient, and less expensive approach would be to use an *on-line* or recursive method. Here, at time n , we use the estimate of the parameters at time $n - 1$ and the observation at time n to update the estimated parameters at time point n . While there is an extensive literature on the recursive estimation of linear stochastic systems, see e.g. the books Ljung and Söderström [64] or Benveniste et. al. [7], the recursive estimation of GARCH processes has attracted little attention until recently.

A recursive estimation method for GARCH processes, supported only by empirical evidence, is developed by Kierkegaard et al. [58]. A recursive method for estimating the parameters of an ARCH process has been presented in Dahlhaus and Subba Rao [25]. For the recursive identification of both ARCH and GARCH models Aknouche and Guerbyenne [1] propose two algorithms. They are based on a suitable transformation of the dynamics that results in an AR or ARMA dynamics in some auxiliary variables, with a driving noise the conditional variance of which is time dependent. Then a weighted least squares or a weighted extended least squares method, see Ljung and Söderström [64] with adaptively chosen weights is proposed. Although this is an elegant approach to the problem, the technical conditions under which the results are valid are not fully specified. Notably, the controversial "boundedness condition" is not discussed, and no remedy, such as resetting, is proposed. In the following we propose a recursive or on-line method for estimating the parameters of a GARCH(r, s) model which is based on the likelihood function constructed by the off-line estimation.

For the solution of the general estimation problem (4.4) the following stochastic approximation procedure is proposed: starting with some initial condition $\theta_0 \in K_0$ we define recursively

$$\theta_n = \theta_{n-1} - \frac{1}{n} \frac{\sigma_{\theta,n}}{\sigma_n} \left(1 - \frac{y_n^2}{\sigma_n^2} \right), \quad (4.11)$$

where σ_n and $\sigma_{\theta,n}$ denote the on-line estimates of $\bar{\sigma}_n(\theta_{n-1})$ and $\bar{\sigma}_{\theta,n}(\theta_{n-1})$, respectively. Thus, at time n , the volatility process σ is generated via the feedback equation

$$[D_{n-1}(z^{-1})(\sigma^2 - \gamma_{n-1})]_n = [C_{n-1}(z^{-1})(y^2 - \gamma_{n-1})]_n$$

with $D_{n-1} = D(z^{-1}, \theta_{n-1})$, and similarly for C_{n-1} .

The convergence properties of the above stochastic gradient methods can be improved, and the analysis can be simplified by using a stochastic Newton method. Recall, that we have

$$R^* = \frac{\partial^2}{\partial \theta^2} W(\theta, \theta^*)|_{\theta=\theta^*} = \lim_{n \rightarrow \infty} 2 \mathbb{E} \frac{\bar{\sigma}_{\theta,n}(\theta) \bar{\sigma}_{\theta,n}(\theta)^T}{\bar{\sigma}_n^2(\theta)}|_{\theta=\theta^*}.$$

Thus the *stochastic Newton* method would read:

$$\theta_n = \theta_{n-1} - \frac{1}{n} R_{n-1}^{-1} \frac{\sigma_{\theta,n}}{\sigma_n} \left(1 - \frac{y_n^2}{\sigma_n^2} \right), \quad (4.12)$$

$$R_n = R_{n-1} + \frac{1}{n} \left(2 \frac{\sigma_{\theta,n} \sigma_{\theta,n}^T}{\sigma_n^2} - R_{n-1} \right). \quad (4.13)$$

The analysis of algorithm (4.11) presented in this section, is based on the BMP-theory, in which a methodology is based on the theory of Markov processes, and its modification by a resetting mechanism, given in Gerencsér and Mátyás [43]. These will be summarized in the next chapter. The analysis is equally applicable to (4.12)-(4.13).

In the following we lay the foundations for the analysis of the proposed recursive algorithm (4.11), modified with resetting. While the methodology of the BMP-scheme is based on the theory of Markov processes, the analysis of our algorithm requires the state space representation of GARCH processes, appropriately modified by the estimated process $(\bar{\sigma}_n(\theta))$. To see this let us extend the state vector X_n^* by

$$\bar{Z}_n(\theta) = (\bar{\sigma}_n^2(\theta), \dots, \bar{\sigma}_{n-s+1}^2(\theta))^T.$$

Then it is easy to see that $\bar{Z}_n(\theta)$ follows the dynamics

$$\bar{Z}_{n+1}(\theta) = B(\theta) \begin{pmatrix} Y_n \\ \bar{Z}_n(\theta) \end{pmatrix} + v(\theta), \quad n \geq 0, \quad (4.14)$$

with

$$B(\theta) = \begin{pmatrix} \eta & \xi \\ 0 & \bar{S} \end{pmatrix} \quad \text{and} \quad v(\theta) = \begin{pmatrix} \alpha_0 \\ 0 \end{pmatrix}, \quad (4.15)$$

where \bar{S} is the shift matrix having 1-s in the sub-diagonal, and 0-s elsewhere,

$$\eta = (\alpha_1, \dots, \alpha_r), \quad \xi = (\beta_1, \dots, \beta_s),$$

and

$$Y_n = (y_n^2, \dots, y_{n-r+1}^2)^T.$$

Thus the extended state-vector

$$\bar{X}_n^e(\theta) = \begin{pmatrix} X_n^* \\ \bar{Z}_n(\theta) \end{pmatrix} \quad (4.16)$$

will follow a linear dynamics with a state-transition matrix of the following triangular structure:

$$A_n^e(\theta) = \begin{pmatrix} A_n^* & 0 \\ x & \xi \\ x & \bar{S} \end{pmatrix}$$

and with input

$$v_n^e(\theta) = \begin{pmatrix} u_n^* \\ v(\theta) \end{pmatrix}.$$

Note that the $(2, 2)$ block of $A_n^e(\theta)$, i.e.

$$\begin{pmatrix} x & \xi \\ x & \bar{S} \end{pmatrix}$$

is stable due to the assumed stability of $D(z^{-1})$.

It is easy to see that $(\bar{X}_n^e(\theta))$ is a (parameter-dependent) Markov process. While the updating function in the algorithm (4.11) contains the derivative of the process $\bar{\sigma}_n^2(\theta)$ with respect to θ , the state vector $\bar{X}_n^e(\theta)$ will be further extended by its derivative with respect to θ , denoted by $\bar{X}_{\theta,n}^e(\theta)$. Differentiating the state-equation for $\bar{X}_n^e(\theta)$ with respect to θ_i , and then collecting these equation for all i -s, we get that $\bar{X}_{\theta,n}^e(\theta)$ follows a linear dynamics with state transition matrix

$$\bar{A}_n^e(\theta) = \text{diag}(A_n^e(\theta), \dots, A_n^e(\theta)).$$

Defining the extended state-vector $\bar{\psi}_n(\theta)$ as

$$\bar{\psi}_n(\theta) = \begin{pmatrix} \bar{X}_n^e(\theta) \\ \bar{X}_{\theta,n}^e(\theta) \end{pmatrix} \quad (4.17)$$

we get that $\bar{\psi}_n(\theta)$ follows a linear dynamics

$$\bar{\psi}_{n+1}(\theta) = P_{n+1}(\theta)\bar{\psi}_n(\theta) + w_n(\theta) \quad (4.18)$$

with a *block-triangular* state transition matrix

$$P_n(\theta) = \begin{pmatrix} A_n^e(\theta) & 0 \\ x & \tilde{A}_n^e(\theta) \end{pmatrix},$$

and $w_n(\theta) = (v_n^e(\theta), v_{\theta,n}^e(\theta))^T$.

It is obvious that $\bar{\psi}_n(\theta)$ is (a parameter dependent) Markov process. Since the asymptotic estimation problem (4.7) can be formulated in terms of $\bar{\psi}_n(\theta)$, the main step in the analysis of the convergence properties of (4.11) is to show that the BMP-theory is applicable to the system (4.18), where the state matrix of the system is *block-triangular*. The foundations of the applicability of the BMP-theory to system (4.18) will be presented in Chapter 6, while a rigorous convergence analysis of the proposed algorithm (4.11) will be given in Chapter 7.

Chapter 5

The general recursive estimation scheme

The first section of this chapter summarizes the basic notions and conditions of the general theory of recursive estimation as presented in Benveniste et al. [7], Chapter 2, Part II. We refer to this setup as BMP scheme, for short. Section 5.2 is devoted to the technical conditions of the BMP theory, and Section 5.3 introduces a special resetting technique developed by Gerencsér and Mátyás [43] which improves the convergence property of the proposed algorithm. Finally, the main result of the chapter on the almost sure convergence of the proposed algorithm is established in Section 5.4.

5.1 The BMP scheme

Following Benveniste et al. [7], we formulate the following general problem. Let (Ω, \mathcal{F}, P) be a probability space. Let $(\bar{X}_n(\theta))$, with $\theta \in D \subset \mathbb{R}^d$, with D being a measurable open set in \mathbb{R}^d , be an \mathbb{R}^k -valued Markov-chain over (Ω, \mathcal{F}, P) with transition kernel $\Pi_\theta(x, A)$, having a unique invariant measure μ_θ . Thus for any $x \in \mathbb{R}^k$ and A , where A is a Borel set of \mathbb{R}^k , $\Pi_\theta(x, A)$ is the probability of moving from x to A in one step. The initial state $\bar{X}_0(\theta)$ is assumed to have distribution μ_θ . Let H be a mapping from $D \times \mathbb{R}^k$ to \mathbb{R}^d . Then the basic estimation problem

of the BMP-theory is to solve the equation

$$E_{\mu_\theta} H(\theta, \bar{X}_n(\theta)) = 0,$$

using observed values of $H(\theta, \bar{X}_n(\theta))$, or their computable approximations. We assume that a unique solution $\theta^* \in D$ exists. For the solution of the above problem the following stochastic approximation procedure is proposed: starting with some initial condition $\theta_0 = \xi$ define recursively

$$\begin{aligned} \theta_n &= \theta_{n-1} + \frac{1}{n} H(\theta_{n-1}, X_n), \\ X_0 &= x_0, \end{aligned} \tag{5.1}$$

where $x_0 \in \mathbb{R}^k$ is a possibly random initial state, and X_n is a non-homogeneous Markov chain defined by

$$P(X_{n+1} \in A | \mathcal{F}_n) = \Pi_{\theta_n}(X_n, A).$$

Here \mathcal{F}_n is the σ -field of events generated by the random variables X_0, \dots, X_n , and A is any Borel subset of \mathbb{R}^k .

Convergence of stochastic approximation (SA) procedures (5.1) given by the BMP theory has been studied in much detail under various sets of assumptions, see Benveniste et al. [7] and Delyon [28]. A similar SA method with mixing rather than Markovian state dynamics has been studied in Gerencsér [41, 42].

A key tool in the analysis of the above method is a so-called ordinary differential equation (ODE) method, see below for details, in which the partial sums of the correction terms in (5.1) are approximated by the solution of an ordinary differential equation. These partial sums are first approximated by a suitable martingale using a standard device in the theory of a Markov chains, namely the *Poisson equation*. This is defined by

$$(I - \Pi_\theta)u = g$$

to be solved for u for a possibly large class of functions g , satisfying $E_{\mu_\theta} g = 0$. A major observation of the BMP theory is that existence and uniqueness of

the solution for a relatively small class of functions g , to be denoted by $Li(q)$, implies existence and uniqueness for a much larger class. The main tool in proving the latter result is the verification that $(\bar{X}_n(\theta))$ is geometrically ergodic for an appropriate class of functions.

5.2 Basic assumptions

In this section we summarize the technical conditions needed in BMP theory.

Let V be a real-valued, measurable function on the state space of the Markov chain (X_n) satisfying $V(x) \geq 1$ for all $x \in \mathbb{R}^k$. Let L_V^∞ denote the set of measurable functions g such that

$$\|g\|_V = \text{ess sup } |g(x)/V(x)| < \infty.$$

Suppose that the Markov chain with transition kernel Π has a unique invariant distribution μ . Recall that, using now the special notations introduced above, the Markov chain is said to be *geometric V -ergodic* if there exist constants $0 < \gamma < 1$ and $K > 0$ such that for all $x \in \mathbb{R}^k$, and all $n \geq 0$ we have

$$\sup_{g \in L_V^\infty} |\Pi_\theta^n g(x) - \mathbb{E}_{\mu_\theta} g| / \|g\|_V \leq K \gamma^n V(x).$$

Note that geometric V -ergodicity implies for all $x, x' \in \mathbb{R}^k$, and all $n \geq 0$

$$|\Pi_\theta^n g(x) - \Pi_\theta^n g(x')| \leq K \gamma^n \|g\|_V (V(x) + V(x')),$$

i.e., the Markov chain forgets its initial condition exponentially fast. The verification of the latter property, the geometric rate of decay of $(\Pi_\theta^n g(x) - \Pi_\theta^n g(x'))$, may indeed be the first step in proving geometric V -ergodicity, see Benveniste et al. [7].

A key innovation of the BMP theory is that the above notion of geometric V -ergodicity is relaxed so that the class of test functions L_V^∞ is replaced by a much more convenient class of test functions, the so-called class $Li(q)$. The starting point is to consider $V(x) = 1 + |x|^q$ with some $q \geq 0$. Define for

measurable real-valued functions g on \mathbb{R}^k the norms

$$\|g\|_q := \text{ess sup}_x \frac{|g(x)|}{1 + |x|^q},$$

and

$$\|\Delta g\|_q = \sup_{x_1 \neq x_2} \frac{|g(x_1) - g(x_2)|}{|x_1 - x_2|(1 + |x_1|^q + |x_2|^q)},$$

and then define the class of test functions as

$$Li(q) = \{ g : \|\Delta g\|_q < +\infty \}.$$

The convenience of BMP theory is that the verification of the exponential decay of $|\Pi_\theta^n g(x) - \Pi_\theta^n g(x')|$ is required only for $g \in Li(q)$. On the other hand, the implied upper bounds for $|\Pi_\theta^n g(x) - E_{\mu_\theta} g|$, will be obtained for a much larger class of functions:

$$C(q+1) = \{ g : g \text{ is continuous and } \|g\|_{q+1} < +\infty \}.$$

It is directly seen that $Li(q) \subseteq C(q+1)$ for any $q \geq 0$. Condition 5.2.1 below addresses exponential forgetting of the Markov transition kernels Π_θ in a technically convenient way:

Condition 5.2.1 *There exists a $q \geq 1$ such that for any compact subset $Q \subset D$ there exist constants K and $0 < \rho < 1$ such that for all $g \in Li(q)$, any $\theta \in Q$ and $x, x' \in \mathbb{R}^k$:*

$$|\Pi_\theta^n g(x) - \Pi_\theta^n g(x')| \leq K \|\Delta g\|_q \rho^n |x - x'| (1 + |x|^q + |x'|^q).$$

Note that in Gerencsér and Mátyás [43] the condition above was required to hold for all q , for the sake of convenience. However, for the present application this would be a much too strong condition. The next condition requires essentially that $\bar{X}_n(\theta)$ is L_{q+1} -bounded.

Condition 5.2.2 *With the q given in Condition 5.2.1 we have: for any compact*

subset $Q \subset D$ there exists a constant K such that for any $\theta \in Q$ and all $x \in \mathbb{R}^k$:

$$\int \Pi_\theta(x, dy)(1 + |y|^{q+1}) \leq K(1 + |x|^{q+1}).$$

Assume furthermore, that with the same $q \geq 1$ for any compact subset $Q \subset D$ there exist $r \in \mathbb{N}$, $0 < \alpha < 1$ and $\beta \in \mathbb{R}$ such that for any $\theta \in Q$ and all $x \in \mathbb{R}^k$:

$$\int \Pi_\theta^r(x, dy)|y|^{q+1} \leq \alpha|x|^{q+1} + \beta.$$

It is easy to show that Condition 5.2.2 implies the existence of a finite constant K' such that for any $\theta \in Q$, all $x \in \mathbb{R}^k$ and any $n \geq 0$

$$\int \Pi_\theta^n(x, dy)(1 + |y|^{q+1}) \leq K'(1 + |x|^{q+1}).$$

We can thus, indeed, say that $\bar{X}_n(\theta)$ is L_{q+1} -bounded.

To ensure an appropriate regularity of the solution of the Poisson equation, denoted by $\nu_\theta(x)$, with respect to θ , we need to impose some regularity conditions both on Π_θ^n and on $H(\theta, x)$, the latter being fairly straightforward.

Condition 5.2.3 *With the q given in Condition 5.2.1 we have: for any compact subset $Q \subset D$ there exists a constant K such that for all $g \in Li(q)$, any $\theta, \theta' \in Q$ and $x \in \mathbb{R}^k$*

$$|\Pi_\theta^n g(x) - \Pi_{\theta'}^n g(x)| \leq K \|\Delta g\|_q |\theta - \theta'| (1 + |x|^{q+1}).$$

Thus the kernels Π_θ^n are in a certain sense Lipschitz-continuous, uniformly in n , with respect to the parameter θ , when applied to the set of test functions $Li(q)$. Note that the prescribed upper bound is independent of n rather than decaying with exponential rate. This is not a mistake: there is no reason to expect that the left hand side decays with n .

As for the updating function $H(\theta, x)$, we need to impose certain growth conditions, along the lines of Benveniste et al. [7]. The synchronization of the exponent p showing up below with the exponent q is obtained by a careful analysis of Benveniste et al. [7], leading to the following:

Condition 5.2.4 For some $p < q/2 - 1$, with q being given in Condition 5.2.1, we have: for any compact subset $Q \subset D$ there exists a constant K depending only on Q satisfying for all $\theta, \theta' \in Q$ and $x \in \mathbb{R}^k$:

$$\begin{aligned} \|\Delta H(\theta, \cdot)\|_p &\leq K \\ |H(\theta, x)| &\leq K(1 + |x|^{p+1}). \end{aligned}$$

Moreover,

$$|H(\theta, x) - H(\theta', x)| \leq K|\theta - \theta'|(1 + |x|^{p+1}).$$

Remark 5.2.1 A great advantage of the theory developed in Benveniste et al. [7] is that it can handle updating functions $H(\theta, x)$ that are *discontinuous* in the variable θ provided the kernels Π_θ have a sufficiently "smoothing effect", i.e. averaging H with respect to x , yielding h , will smooth out eventual discontinuities. This can be done by requiring the above condition, ultimately, for the smoothed function $\Pi_\theta H_\theta$. To see this consider the Poisson equation

$$(I - \Pi_\theta)\nu(x) = H(\theta, x) - h(\theta). \quad (5.2)$$

The key observation is that instead of requiring the Lipschitz-regularity of the solutions $\nu_\theta(x)$ it suffices to require the regularity of the functions $\Pi_\theta \nu_\theta(x)$; this, in turn, is satisfied if the averaged function $\Pi_\theta H(\theta, x)$ is smooth in θ and the solutions $w = w_\theta$ to the "smoothed" Poisson equation

$$(I - \Pi_\theta)w(x) = \Pi_\theta H(\theta, x) - h(\theta) \quad (5.3)$$

are regular with respect to θ . To see the connection more clearly, note that if ν_θ is the solution to the original Poisson equation (5.2), then

$$w_\theta(x) = \Pi_\theta \nu_\theta(x)$$

is the solution of the smoothed Poisson equation (5.3):

$$(I - \Pi_\theta)w_\theta(x) = \Pi_\theta(I - \Pi_\theta)\nu_\theta(x) = \Pi_\theta H(\theta, x) - h(\theta).$$

Conversely, if $w_\theta(x)$ is a solution of the smoothed Poisson equation (5.3) then

$$\nu_\theta(x) = w_\theta(x) + H(\theta, x) - h(\theta)$$

is a solution of the Poisson equation (5.2). Indeed,

$$\begin{aligned} (I - \Pi_\theta) \nu_\theta(x) &= (\Pi_\theta H(\theta, x) - h(\theta)) + (H(\theta, x) - \Pi_\theta H(\theta, x)) = \\ &= H(\theta, x) - h(\theta). \end{aligned}$$

The observation regarding the regularity of the functions $\Pi_\theta \nu_\theta$ now follows since

$$\begin{aligned} \Pi_\theta \nu_\theta(x) - \Pi_{\theta'} \nu_{\theta'}(x) &= \Pi_\theta w_\theta(x) - \Pi_{\theta'} w_{\theta'}(x) + \\ &+ \Pi_\theta H(\theta, x) - \Pi_{\theta'} H(\theta', x) + h(\theta) - h(\theta'). \end{aligned}$$

The above considerations motivate the following modified assumption of Benveniste et al. [7] concerning the updating function $H(\theta, x)$.

Condition 5.2.5 *For some $p < q/2 - 1$, with q being given in Condition 5.2.1 and for any compact subset $Q \subset D$ there exists a constant K depending only on Q satisfying for all $\theta, \theta' \in Q$ and any $x \in \mathbb{R}^k$:*

$$\begin{aligned} \|\Delta \Pi_\theta H(\theta, \cdot)\|_p &\leq K \\ |H(\theta, x)| &\leq K(1 + |x|^{p+1}) \\ |\Pi_\theta H(\theta, x) - \Pi_{\theta'} H(\theta', x)| &\leq K|\theta - \theta'|(1 + |x|^{p+1}). \end{aligned}$$

Note that the final inequality is a formal expression of the smoothing effect of Π_θ when applied to $H(\theta, \cdot)$.

Theorem 5, p. 259 of Benveniste et al. [7] now yields the following basic result.

Theorem 5.2.1 (Benveniste, Metivier and Priouret [7]) *Assume that the Markov kernels Π_θ satisfy Conditions 5.2.2 - 5.2.3 and H satisfies Condition 5.2.4. Then we have with*

$$h(\theta) = \Gamma_\theta H(\theta, \cdot) = \int H(\theta, x) \mu_\theta(dx)$$

(i) For all $\theta \in D$ the Poisson equation

$$(I - \Pi_\theta) \nu(x) = H(\theta, x) - h(\theta)$$

has a unique solution $\nu_\theta = \nu_\theta(x)$.

(ii) h is locally Lipschitz-continuous on D .

(iii) For any compact subset $Q \subset D$ there exists a constant C such that for all $\theta \in Q$:

$$|\nu_\theta(x)| \leq C(1 + |x|^{p+1}). \quad (5.4)$$

(iv) For any compact subset $Q \subset D$ and any $0 < \lambda < 1$ there exists a constant C_λ such that for all $\theta, \theta' \in Q$ and any $x \in \mathbb{R}^k$:

$$|\Pi_\theta \nu_\theta(x) - \Pi_{\theta'} \nu_{\theta'}(x)| \leq C_\lambda |\theta - \theta'|^\lambda (1 + |x|^{p+1}).$$

Note that the $\nu_\theta(x)$ is only Hölder-continuous with respect to θ as opposed to the Lipschitz-continuity of $\Pi_\theta H(\theta, x)$.

Remark 5.2.2 By virtue of Remark 5.2.1 and Theorem 5.2.1 it can be easily seen that, under Condition 5.2.5, the solution of the smoothed Poisson equation $w_\theta(x) = \Pi_\theta \nu_\theta(x)$ satisfies (5.4).

Remark 5.2.3 It can be realized that most of the effort of the BMP theory is directed towards the study of the solution of the Poisson equation. This is a very convenient tool for convergence analysis developed for the study of stochastic approximations and adaptive algorithms. A classical well-known result of the theory of the Poisson equation is as follows. Suppose that we are interested in the behavior of

$$\frac{1}{N} \sum_{n=1}^N H(\theta, X_n(\theta)),$$

where $X_n(\theta)$ is a parametric Markov process with values in \mathbb{R}^k , transition kernel

Π_θ , and invariant measure π . Suppose that the Poisson equation

$$(I - \Pi_\theta)\nu(x) = H(\theta, x) - h(\theta)$$

has a solution $\nu(x)$ on \mathbb{R}^k . From the application of the Poisson equation it follows that

$$\frac{1}{N} \sum_{n=1}^N (H(\theta, X_n(\theta)) - h(\theta)) = \frac{1}{n} \sum_{n=1}^N (I - \Pi_\theta)\nu(X_n(\theta)),$$

which, rearranging terms, results in

$$\begin{aligned} \frac{1}{N} \sum_{n=1}^N (H(\theta, X_n(\theta)) - h(\theta)) &= \frac{1}{N} \sum_{n=1}^N (\nu(X_n(\theta)) - \Pi_\theta \nu(X_{n-1}(\theta))) + \\ &\quad + \frac{1}{N} (\nu(X_1(\theta)) - \Pi_\theta \nu(X_N(\theta))). \end{aligned}$$

Since the terms in the sum are bounded martingale differences, and ν_θ and $\Pi_\theta \nu$ are bounded, hence

$$\frac{1}{N} \sum_{n=1}^N (H(\theta, X_n(\theta)) - h(\theta))$$

goes to zero with N , from the law of large numbers for martingales, see Hall and Heyde [51]. This observation will be the main tool by providing strong approximation result on the error term of the off-line quasi-maximum likelihood estimator.

5.3 Resetting

Convergence of the estimator sequence (θ_n) with probability *strictly less than 1* has been proved in Benveniste et al. [7], in Theorem 13, p. 236 of Part II. A critical issue in the analysis of the BMP scheme is that θ_n may eventually leave its domain of definition. This is sometimes referred to as the boundedness problem. A simple remedy for this is to restrict the process to D , or to a compact truncation domain $D_0 \subset D$ containing θ^* in its interior, by stopping the process if θ_n would leave D_0 . This situation is getting even worse, when an

additional criterion for stopping is introduced: namely, when the process will be stopped, if the difference between two successive estimators exceeds a fixed threshold. A set of technical conditions and rigorous analysis of the effect of resettings has been first given by Gerencsér [41].

A BMP-scheme modified by an appropriate time-varying resetting mechanism has been proposed and analyzed by Delyon [28]. Gerencsér and Mátyás [43] use a fixed, fairly arbitrary truncation domain and a fixed threshold to prevent the estimates from making large jumps, and in addition their conditions are more realistic, in fact almost identical with the conditions of Benveniste et al. [7]. In the following we present the modified stochastic approximation algorithm introduced by Gerencsér and Mátyás [43], using a suitable resetting mechanism, which is shown to converge with *probability 1* to θ^* , under reasonable technical conditions.

Since the updating function H is defined on $D \times \mathbb{R}^k$, the SA algorithm (5.1) makes sense only if $\theta_{n-1} \in D$. Hence we will require that the estimator lie in a compact truncation domain $D_0 \subseteq D$, with properties specified below. In addition we would like to prevent the estimates from making large jumps. Therefore we have to modify our algorithm to enforce these boundedness conditions. Let us choose a fixed small number $\varepsilon > 0$. Let $\tau_0 \equiv 0$ and for $i \geq 1$ define recursively the stopping times

$$\tau_i := \min\{\tau_i^e, \tau_i^j\},$$

where

$$\tau_i^e = \inf\{k > \tau_{i-1} : \theta_k \notin D_0\}$$

$$\tau_i^j = \inf\{k > \tau_{i-1} : |\theta_k - \theta_{k-1}| > \varepsilon\}.$$

τ_i is thus the first time after τ_{i-1} at which the algorithm either leaves D_0 or a jump of magnitude at least ε occurs. At τ_i we re-initialize both the parameter value and the state vector of the algorithm: we set

$$\theta_{\tau_i} := \xi_0 \quad \text{and} \quad X_{\tau_i} := x_0.$$

To formalize the resetting procedure let θ_{i-} denote the value of θ computed at

time i by (5.1) and define the set

$$B_i = \{\omega \mid \theta_{i-} \notin D_0 \text{ or } |\theta_{i-} - \theta_{i-1}| > \varepsilon\}.$$

Then *the recursive algorithm with resetting* is defined as

$$\theta_i = \theta_{i-1} + (1 - \chi_{B_i}) \frac{1}{i} H(\theta_{i-1}, X_i) + \chi_{B_i} (\xi_0 - \theta_{i-1}), \quad (5.5)$$

with $\theta_0 = \xi_0$, where X_i follows the dynamics

$$X_i = (1 - \chi_{B_i}) f(X_{i-1}, \theta_{i-1}, U_i) + \chi_{B_i} x_0, \quad X_0 = x_0,$$

realizing the Markovian dynamics, with a Borel-measurable mapping f from $\mathbb{R}^k \times D \times [0, 1]$ to \mathbb{R}^k , and a random variable U uniformly distributed on $[0, 1]$.

5.4 Convergence

The convergence of the sequence (θ_n) is analyzed using a so-called ODE-method. The ODE-method is stated in various forms in Kushner and Clark [59], Ljung and Söderström [64] and Benveniste et al. [7]. For SA procedures with resetting see Gerencsér [41, 42]. The ODE method indicates that the (θ_n) is closely related to the solution of the associated ODE

$$\frac{d}{ds} \bar{\theta}_s = h(\bar{\theta}_s), \quad \bar{\theta}_0 = \xi. \quad (5.6)$$

Let $\bar{\theta}(t, s, \xi)$ denote the general solution. To ensure convergence of (θ_n) we have to require asymptotic stability of the associated ODE. To ensure that the parameter sequence (θ_n) is not bounced back and forth by resetting we need to impose some conditions on the shape of the truncation domain D_0 and on the position of the initial value ξ relative to D_0 . Condition 5.4.1 below is taken from Gerencsér and Mátyás [43].

Condition 5.4.1 *Let $D_0 \subset D$ be a compact truncation domain. Assume that for any $\xi \in D_0$, $\bar{\theta}(t, 0, \xi) \in D$ is defined for any $t \geq 0$, and for some $\theta^* \in \text{int } D_0$*

we have

$$\lim_{t \rightarrow \infty} \bar{\theta}(t, 0, \xi) = \theta^*$$

for any initial value $\xi \in D_0$. Assume furthermore, that we have an initial estimate ξ_0 such that for all $t \geq 0$ we have

$$\bar{\theta}(t, 0, \xi_0) \in \text{int}D_0.$$

The theorem below is essentially given in Gerencsér and Mátyás [43]:

Theorem 5.4.1 (Gerencsér and Mátyás [43]) *Consider algorithm (5.5) and assume that Condition 5.2.2-5.4.1 hold. Let ε , the limiting rate of θ_n , be sufficiently small. Then*

$$\lim_n \theta_n = \theta^* \quad \text{w.p.1,}$$

and also in L_q , for $q \geq 1$ given in Condition 5.2.1, with rate

$$E^{1/q} |\theta_n - \theta^*|^q = O(n^{-(\alpha \wedge \frac{1}{2})}),$$

where $-\alpha < 0$ is the Lyapunov-exponent of the associated ODE.

Remark 5.4.1 The Lyapunov-exponent of the associated ODE is given by the maximum of the real-parts of the eigenvalues of the Jacobian-matrix of the right hand side at $\theta = \theta^*$. An analogous theorem is valid for the stochastic Newton method. In this case $\alpha = -1$, and the convergence rate in L_q is $O(n^{-\frac{1}{2}})$.

Remark 5.4.2 Note, that the asymptotic stability of the associated ODE (5.6) was guaranteed in Benveniste et al. [7] by the existence of a global Lyapunov function defined on the *whole* domain D which is assumed to be an invariant domain for the ODE (see p. 233 of [7]):

Condition 5.4.2 *There exists $\theta^* \in D$ and a positive function U of class C^2 with domain of definition D such that*

$$(i) \quad U(\theta) > U(\theta^*) = 0 \text{ for all } \theta \in D, \theta \neq \theta^*,$$

$$(ii) \quad U'(\theta)h(\theta) < 0 \text{ for all } \theta \in D, \theta \neq \theta^*,$$

(iii) $U(\theta) \rightarrow \infty$ if $\theta \rightarrow \partial D$ or $|\theta| \rightarrow +\infty$.

Note that Condition 5.4.2 automatically entails Condition 5.4.1.

Chapter 6

Stability properties of block-triangular stationary random matrices

This chapter deals with the stability properties of block-triangular stationary random matrices. Section 6.1 presents the technical conditions and the main results related to the top-Lyapunov exponent and the L_q -stability of certain linear stochastic systems with block-triangular state matrices. Section 6.2 is devoted to the proofs of the results of the first section and a nice subsection on the delicacy of the proof of Theorem 6.1.1. All of the results of this chapter are based on the articles Gerencsér, Michaletzky and Orlovits [44] and Gerencsér and Orlovits [45]. The major area of the application of the results of this chapter will be the analysis of the recursive estimation method for GARCH processes appearing below in Chapter 7.

6.1 Introduction and the main results

Consider again a linear stochastic system given by equation of the form

$$X_{n+1} = A_{n+1}X_n + u_{n+1}, \quad -\infty < n < +\infty, \quad (6.1)$$

where $X_n \in \mathbb{R}^d$, A_n is a random matrix in $\mathbb{R}^{d \times d}$ and u_n is a random vector in \mathbb{R}^d . Throughout this chapter we will assume again that Condition 3.2.1 holds, that is (A_n, u_n) is a jointly strictly stationary, ergodic sequence of $d \times (d+1)$ random matrices over some probability space (Ω, \mathcal{F}, P) . Let us recall from Chapter 3 that to formulate a sufficient condition for the existence of a strictly stationary and ergodic solution of (6.1) we need the concept of a top-Lyapunov exponent defined by

$$\lambda = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \log \|A_n \dots A_1\|$$

with $\mathcal{A} = (A_n)$ be as above such that

$$\mathbb{E} \log^+ \|A_n\| < +\infty. \quad (6.2)$$

A useful corollary of the Furstenberg-Kesten Theorem is that for all $\varepsilon > 0$ there exists a finite random variable $C_A(\omega)$ such that

$$\|A_n \dots A_1\| \leq C_A(\omega) e^{n(\lambda + \varepsilon)}. \quad (6.3)$$

We have seen in Chapter 3 that if (A_n, u_n) satisfies Condition 3.2.1, (6.2) holds for (A_n) and

$$\mathbb{E} \log^+ |u_n| < +\infty, \quad (6.4)$$

then $\lambda(\mathcal{A}) < 0$ implies that (6.1) has a unique strictly stationary, causal solution given by

$$X_n^* = u_n + \sum_{k=1}^{\infty} A_n A_{n-1} \dots A_{n-k+1} u_{n-k}. \quad (6.5)$$

Furthermore, solving (6.1) with any initial condition X_0 forward in time we get that for any $\varepsilon > 0$ we have

$$X_n - X_n^* = O(e^{(\lambda + \varepsilon)n}) \quad \text{w.p.1.} \quad (6.6)$$

Assume now that the matrices (A_n, u_n) depend on a scalar parameter $\theta \in D \subset \mathbb{R}$, where D is an open set, and assume that $(A_n(\theta), u_n(\theta))$ are C^1 functions of θ for all $\omega \in \Omega$. Suppose that Condition 3.2.1 holds, and $\lambda(\mathcal{A}) < 0$ for all $\theta \in D$. The unique strictly stationary, causal solution of (6.1) will be denoted by $(X_n(\theta))$. In system identification we typically need to differentiate a parameter-dependent state-vector. It is not obvious at all that $X_n(\theta)$ is differentiable for all ω or for almost all $\omega \in \Omega$.

Note that we do not have this problem if we solve the recursion for $X_n(\theta)$ starting with $n = 0$. Let $X_{\theta,n}(\theta)$ denote the derivative of $X_n(\theta)$ with respect to the parameter vector θ . For the sake of simplicity we drop the dependence on θ . Carrying out formal differentiation and assuming that θ is scalar we get for the extended state vector $(X_n, X_{\theta,n})$ the linear state-space equation

$$\begin{aligned} X_{n+1} &= A_{n+1}X_n + u_{n+1}, \\ X_{\theta,n+1} &= A_{\theta,n+1}X_n + A_{n+1}X_{\theta,n} + u_{\theta,n+1}, \end{aligned} \tag{6.7}$$

where $A_{\theta,n}$ denote the derivative of A_n with respect to θ . Thus the state transition matrix will be

$$\bar{A}_n = \begin{pmatrix} A_n & 0 \\ A_{\theta,n} & A_n \end{pmatrix}.$$

First we may then ask if $\lambda(\mathcal{A}) < 0$ implies $\lambda(\bar{\mathcal{A}}) < 0$, and if an approximation similar to (6.6) is possible for the solutions of the extended system (6.7). A positive answer to this problem has been given in Mikosch and Straumann [70] under the condition that the sequence (A_n) is i.i.d., and for some $s > 0$

$$E\|\bar{A}_1\|^s < +\infty.$$

The proof of Mikosch and Straumann [70] follows a route completely different from what will be developed in Section 6.2, and apparently can not be generalized to cover the general case to be discussed here. For the sake of completeness we present the main tool that is used in the proof of Mikosch and Straumann [70]:

Lemma 6.1.1 (Mikosch and Straumann [70]) *Let $\mathcal{A} = (A_n)$ be an i.i.d. sequence of $d \times d$ matrices with $E\|A_n\|^s < \infty$ for some $s > 0$. Then for the associated top-Lyapunov exponent we have $\lambda(\mathcal{A}) < 0$ if and only if there exist $c > 0$, $\bar{s} > 0$ and $0 < \rho < 1$ such that*

$$E\|A_n \dots A_1\|^{\bar{s}} \leq c\rho^n \quad \text{for } n \geq 1.$$

The proof is based on the observation that for any fixed $l \geq 1$ the map $u \mapsto E\|A_l \dots A_1\|^u$ has first derivative with respect to u equal to $E \log \|A_l \dots A_1\|$ at $u = 0$.

The first result of this chapter is a positive answer to this problem in a general setting under significantly *weaker* conditions.

Theorem 6.1.1 *Let*

$$A_n = \begin{pmatrix} A_n^1 & 0 \\ B_n & A_n^2 \end{pmatrix}$$

be a stationary, ergodic sequence of $(d_1 + d_2) \times (d_1 + d_2)$ matrices, satisfying (6.2), with A_n^1 and A_n^2 being square matrices. Then

$$\lambda(\mathcal{A}) = \max(\lambda(\mathcal{A}^1), \lambda(\mathcal{A}^2)).$$

The proof of this theorem is given in the next section. Note, that the above result for i.i.d. sequences follows from the results of Furstenberg and Kifer [38], Lemma 3.6., p. 24. and from Subba Rao [85], Lemma 3.1, p.1161, as well.

A direct consequence of the above theorem is that an approximation similar to (6.6) is possible for the solution of the extended system (6.7).

Corrolary 6.1.1 *Let (A_n, u_n^1) be a stationary, ergodic sequence and assume that (6.2) and (6.4) holds for (A_n) and (u_n) , respectively. Consider a linear stochastic system*

$$X_n^1 = A_n^1 X_{n-1}^1 + u_n^1 \tag{6.8}$$

$$X_n^2 = A_n^2 X_{n-1}^2 + B_n X_{n-1}^1 + u_n^2, \tag{6.9}$$

where $X_n = (X_n^1, X_n^2) \in \mathbb{R}^{d_1+d_2}$, $u_n = (u_n^1, u_n^2)$, and $A_n = \begin{pmatrix} A_n^1 & 0 \\ B_n & A_n^2 \end{pmatrix}$. Then solving (6.8) with any initial condition X_0 forward in time we get that for any $\varepsilon > 0$

$$X_n - X_n^* = O(e^{(\lambda+\varepsilon)n}) \quad \text{w.p.1.} \quad (6.10)$$

Remark 6.1.1 An application and motivating example for the above theorem consider again the random linear stochastic system (6.1) satisfying $\lambda(\mathcal{A}) < 0$. We have seen that condition (6.2) is sufficient for the existence of a stationary and causal solution. We give an example, using Theorem 6.1.1, which shows that this condition might be relaxed. Consider *serially coupled* input-output systems given by the equations, with $-\infty < n < \infty$,

$$X_n^1 = A_n^1 X_{n-1}^1 + u_n^1 \quad (6.11)$$

$$X_n^2 = A_n^2 X_{n-1}^2 + B_n X_{n-1}^1, \quad (6.12)$$

where $u_n^1 \in \mathbb{R}^{d_1}$. Note that there is no exogenous input in (6.12). Defining

$$u_n^2 = B_n X_{n-1}^1, \quad (6.13)$$

we can write (6.12) as

$$X_n^2 = A_n^2 X_{n-1}^2 + u_n^2. \quad (6.14)$$

Note, however, that the validity of (6.4) for u_n^2 can not be guaranteed.

In spite of this (6.14) is well defined under the conditions of the following theorem, a direct consequence of Theorem 6.1.1.

Corrolary 6.1.2 Let (A_n, u_n^1) be a stationary, ergodic sequence, jointly satisfying (6.2) and (6.4). Assume that $\lambda(\mathcal{A}_1) < 0$, $\lambda(\mathcal{A}_2) < 0$. Let (X_n^1) be the unique strictly stationary, causal solution of (6.11). Then the linear stochastic

system (6.14) has a strictly stationary, causal solution given by

$$X_n^2 = B_n X_{n-1}^1 + \sum_{k=1}^{\infty} A_n^2 A_{n-1}^2 \cdots A_{n-k+1}^2 u_{n-k}^2, \quad (6.15)$$

where the right hand side converges almost surely.

The purpose of the next part of this section is to extend the result of Theorem 6.1.1 on the top-Lyapunov exponent of a stationary, ergodic sequence of block-triangular random matrices to the problem of L_q -stability for i.i.d. sequences of block-triangular random matrices.

Consider again the linear stochastic system given by equation (6.1). Assume that Condition 3.2.1 holds, and suppose furthermore that

$$\mathbb{E}|u_n|^q < +\infty \quad \text{for some } q \geq 1.$$

To simplify the discussion let us recall Condition 3.3.1 from Chapter 3, that is (A_n) is an i.i.d. sequence of random matrices, and the two σ -fields

$$\mathcal{F}_{n+}^A = \sigma\{A_i : i > n\}$$

$$\mathcal{F}_{n-}^u = \sigma\{u_i : i \leq n\}$$

are independent for any n .

We have seen in Chapter 3 that under Conditions 3.2.1 and Condition 3.3.1 a sufficient condition for L_q -convergence in (6.5) is that $\rho[\mathbb{E}(A_n^{\otimes q})] < 1$, with q being a positive even integer. Similarly to the case of the top-Lyapunov exponent, we can then ask if L_q -stability is inherited by the extended system. A second, related question if the sufficient condition for L_q -stability, $\rho[\mathbb{E}(A^{\otimes q})] < 1$, with q being a positive even integer, is inherited by the extended system. The answer to the first question is implied directly by Lemma 4.9 of Mikosch and Straumann [70]:

Proposition 6.1.1 (Mikosch and Straumann [70]) *Let*

$$A_n = \begin{pmatrix} A_n^1 & 0 \\ B_n & A_n^2 \end{pmatrix}$$

be an i.i.d. sequence of random $(d_1 + d_2) \times (d_1 + d_2)$ matrices such that $\mathbb{E} \|A_n\|^s < +\infty$ for some $s > 0$. Then

$$\lambda_s(\mathcal{A}) = \max(\lambda_s(\mathcal{A}^1), \lambda_s(\mathcal{A}^2)).$$

The second result of this chapter is a positive answer to the second question, related to the sufficient condition for L_q -stability. Note that in the theorem below q can be any positive integer, even or odd.

Theorem 6.1.2 *Let*

$$A = \begin{pmatrix} A_1 & 0 \\ B & A_2 \end{pmatrix}$$

be a random $(d_1 + d_2) \times (d_1 + d_2)$ matrix in $L_2(\Omega, \mathcal{F}, P)$, with A_1 and A_2 being square matrices. Then

$$\rho[\mathbb{E}(A \otimes A)] = \max\{ \rho[\mathbb{E}(A_1 \otimes A_1)] ; \rho[\mathbb{E}(A_2 \otimes A_2)] \}. \quad (6.16)$$

Similarly, let q be a positive integer, even or odd and let us assume that $A \in L_q(\Omega, \mathcal{F}, P)$. Then

$$\rho[\mathbb{E}(A^{\otimes q})] = \max\{ \rho[\mathbb{E}(A_1^{\otimes q})] ; \rho[\mathbb{E}(A_2^{\otimes q})] \}. \quad (6.17)$$

The proof is purely algebraic, and will be given in the next section. The key tool is a Cauchy-Schwartz inequality stated for tensor-products, given as Lemma 6.2.7 below.

6.2 Proofs

6.2.1 Proof of Theorem 6.1.1

It can be easily seen that the inequality $\lambda(\mathcal{A}) \geq \max(\lambda(\mathcal{A}^1), \lambda(\mathcal{A}^2))$ holds trivially. We will show that

$$\lambda(\mathcal{A}) \leq \max(\lambda(\mathcal{A}^1), \lambda(\mathcal{A}^2))$$

holds as well. It is easy to see that it is sufficient to show that

$$\lambda(\mathcal{A}^1) < 0, \quad \lambda(\mathcal{A}^2) < 0 \quad \text{implies} \quad \lambda(\mathcal{A}) < 0. \quad (6.18)$$

Indeed, for any constant $c > 0$ setting $c\mathcal{A} = (cA_n)$ obviously

$$\lambda(c\mathcal{A}) = \lambda(\mathcal{A}) + \log c.$$

Let c be such that $\lambda(c\mathcal{A}^1) < 0$ and $\lambda(c\mathcal{A}^2) < 0$. Then by (6.18) it follows that $\lambda(c\mathcal{A}) < 0$. In other words, with $\gamma = -\log c$, we have

$$\lambda(\mathcal{A}^1) < \gamma \quad \lambda(\mathcal{A}^2) < \gamma \quad \text{implies} \quad \lambda(\mathcal{A}) < \gamma,$$

from which the claim follows. To prove (6.18) we use the following result which is part of Oseledec's theorem, see Oseledec [75] or Ragunathan [80].

Theorem 6.2.1 *Let $\mathcal{A} = (A_n)$ be a stationary, ergodic sequence of $d \times d$ matrices satisfying (6.2). Then there are random subspaces $S(\omega) \subset \mathbb{R}^d$ of fixed dimension, say, $s < d$, such that S is a measurable function from (Ω, \mathcal{F}) to the Grassmanian manifold of s -dimensional subspaces of \mathbb{R}^d , and there exists a null-set $\Omega_0 \subset \Omega$ such that for $\omega \notin \Omega_0$ and any $x \in \mathbb{R}^d, x \notin S(\omega)$ we have*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log |A_n \dots A_1 x| = \lambda(\mathcal{A}).$$

Remark 6.2.1 The original form of Oseledec's theorem has been stated for non-singular random matrices, see Oseledec [75], but an extension to possibly singular sequences has been given in Ragunathan [80]. Although there seems to be a gap in the arguments of Ragunathan [80], the proof of the above partial result is rigorous.

Corollary 6.2.1 *Under the conditions of Theorem 6.2.1 we have $\lambda(\mathcal{A}) < 0$ if and only if for Lebesgue almost all $x \in \mathbb{R}^d$ we have*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log |A_n \dots A_1 x| < 0 \quad \text{w.p.1.} \quad (6.19)$$

We note in passing that, obviously, the hard part of the above result is to show that (6.19) implies $\lambda(\mathcal{A}) < 0$.

To prove (6.18) we apply Corollary 6.2.1. Thus it is sufficient to show that for Lebesgue almost all $x \in \mathbb{R}^{d_1+d_2}$

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log |A_n \dots A_1 x| < 0 \quad \text{w.p.1.} \quad (6.20)$$

Writing

$$\begin{aligned} x_n^1 &= A_n^1 x_{n-1}^1 & x_0^1 &= x^1 \\ x_n^2 &= B_n x_{n-1}^1 + A_n^2 x_{n-1}^2 & x_0^2 &= x^2 \end{aligned}$$

and solving this recursion we get

$$x_n^1 = A_n^1 \dots A_1^1 x_0^1 \quad (6.21)$$

$$x_n^2 = A_n^2 \dots A_1^2 x_0^2 + \sum_{k=1}^n A_n^2 \dots A_{k+1}^2 B_k A_{k-1}^1 \dots A_1^1 x_0^1. \quad (6.22)$$

To estimate x_n^1 (6.3) is applicable. Thus we get

$$|A_n^1 \dots A_1^1 x_0^1| \leq C_A^1(\omega) e^{n(\lambda^1 + \varepsilon)} |x_0^1| \quad (6.23)$$

with $\lambda^1 = \lambda(\mathcal{A}^1)$. The first term on the right hand side of (6.22) is estimated similarly. To estimate the norm of the second term of (6.22) we apply first the triangle inequality, and then estimate the k -th term. We first estimate $\|A_{k-1}^1 \dots A_1^1\|$ using the inequality (6.3) with $(k-1)$ replacing n . To estimate $\|B_k\|$ we use the following simple well-known lemma:

Lemma 6.2.1 *Let $(\eta_k), k \geq 1$ be a sequence of identically distributed real-valued random variables such that $E \log^+ |\eta_k| < +\infty$. Then for any $\varepsilon > 0$ there exists a finite random variable $C(\omega)$ such that for all $k \geq 1$*

$$|\eta_k| \leq C(\omega) e^{\varepsilon k}.$$

In other words (η_k) is sub-exponential.

Thus we get that for any $\varepsilon > 0$ there exists a finite random variable C_B such that for all $k \geq 1$

$$\|B_k\| \leq C_B e^{\varepsilon k}.$$

Therefore the norm of the second term in (6.22) can be estimated from above by

$$C_A^1 C_B \sum_{k=1}^n \|A_n^2 \dots A_{k+1}^2\| e^{k(\lambda^1 + 2\varepsilon)} |x_0^1|, \quad (6.24)$$

and here $\lambda^1 < 0$. To estimate $\|A_n^2 \dots A_{k+1}^2\|$ we prove a few simple auxiliary results:

Lemma 6.2.2 *Let $(\xi_n), n \geq 1$ be a strictly stationary, ergodic process with $E\xi_n = 0$. Then for all $\varepsilon > 0$ there exists a random constant $c(\omega)$ such that w.p.1*

$$\xi_1 + \dots + \xi_n \leq c(\omega) + n\varepsilon.$$

Proof. The proposition is a direct consequence of the strong law of large numbers, known as Birkhoff's ergodic theorem, for strictly stationary, ergodic processes. ■

Similarly, reversing the time, we would get for a two-sided, strictly stationary, ergodic process that

$$\xi_n + \dots + \xi_{k+1} \leq c_n(\omega) + (n - k)\varepsilon.$$

Unfortunately, the constant $c_n(\omega)$ depends on n . To get an upper bound in which the constant does not depend on n , write

$$\begin{aligned} \xi_n + \dots + \xi_{k+1} &= (\xi_1 + \dots + \xi_n) - (\xi_1 + \dots + \xi_k) \\ &\leq 2 \sup_{1 \leq k \leq n} (\xi_1 + \dots + \xi_k) \leq 2(c(\omega) + n\varepsilon). \end{aligned}$$

Repeating this argument with $\varepsilon/2$ replacing ε we get the following lemma:

Lemma 6.2.3 *Let $(\xi_n), n \geq 1$ be a strictly stationary, ergodic process with $E\xi_n = 0$. Then for all $\varepsilon > 0$ there exists a random constant $c(\omega)$ such that for all $0 \leq k < n$ we have w.p.1*

$$\xi_n + \dots + \xi_{k+1} \leq c(\omega) + n\varepsilon.$$

In the case when $E\xi_n \neq 0$ we have the following result:

Lemma 6.2.4 *Let $(\xi_n), n \geq 1$ be a strictly stationary, ergodic process such that $E\xi_n$ exists and $E\xi_n < \infty$. Then for all finite $\mu \geq E\xi_n$, and for all $\varepsilon > 0$ there exists a random constant $c(\omega)$, such that for all $0 \leq k < n$ we have w.p.1*

$$\xi_n + \dots + \xi_{k+1} \leq c(\omega) + (n - k)\mu + n\varepsilon.$$

Proof. If $E\xi_n$ is finite, then, applying Lemma 6.2.3 for the random variables $\bar{\xi}_n = \xi_n - E\xi_n$, the claim follows for $\mu = E\xi_n$, thus, by monotonicity, it follows also for larger μ -s. If $E\xi_n = -\infty$, then first define $\bar{\xi}_n = \xi_n \vee (-K)$, where K is large enough to ensure that $E\bar{\xi}_n \leq \mu$. Then the first part is applicable to get

$$\bar{\xi}_n + \dots + \bar{\xi}_{k+1} \leq c(\omega) + (n - k)\mu + n\varepsilon \quad \text{w.p.1.}$$

Since $\bar{\xi}_n \geq \xi_n$, we get the claim. ■

We continue the proof of Theorem 6.1.1 by estimating $\|A_n^2 \dots A_{k+1}^2\|$ first in the scalar case, i.e., when A_n^2 is a scalar, say $|A_n^2| = a_n$. Then obviously $\lambda^2 = \lambda(\mathcal{A}^2) = E \log a_n$.

Lemma 6.2.5 *Let $(a_n), n \geq 1$ be a non-negative, strictly stationary, ergodic process such that $\lambda = E \log a_k$ exists, and $\lambda < \infty$. Then for all finite $\mu \geq \lambda$, and for any $\varepsilon > 0$ there exists a random variable $C(\omega)$, such that for all $0 \leq k < n$*

$$a_n \dots a_{k+1} \leq C(\omega) e^{(n-k)\mu} e^{n\varepsilon} \quad \text{w.p.1.} \quad (6.25)$$

A key point in the above statement is that $C(\omega)$ is independent of n .

Proof. Writing $\xi_k = \log a_k$, applying Lemma 6.2.4, then exponentiating the resulting inequality, we get the claim. ■

Using this lemma with $\lambda = \lambda(\mathcal{A}^2) = \lambda^2$ in (6.24), substituting the resulting upper bound into (6.22), and using the arguments following (6.22), we get after some simplifications that for

$$\mu = \max(\lambda^1, \lambda^2) < 0,$$

and for any ε there exists a random variable $C'(\omega)$, such that for all $n \geq 1$ we have

$$|x_n^2| \leq C'(\omega)e^{n(\mu+\varepsilon)},$$

and thus (6.20) follows.

To estimate $\|A_n^2 \dots A_{k+1}^2\|$ in the general case we need a simple observation which states that the subadditive process $\log \|A_n^2 \dots A_{k+1}^2\|$ can be majorated by a scalar valued additive process modulo negligible error with growth rate arbitrary close to $\lambda^2 = \lambda(\mathcal{A}^2) < 0$.

Lemma 6.2.6 *Let $\mathcal{A} = (A_n)$ be a strictly stationary, ergodic sequence of $d \times d$ random matrices such that $E \log^+ \|A_n\| < +\infty$. Then for any $\varepsilon > 0$ there exists a scalar valued, stationary and ergodic process (ξ_n) , and a finite random variable $c(\omega)$, such that $E\xi_n < \lambda(\mathcal{A}) + \varepsilon$, and for any $0 \leq k < n$*

$$\log \|A_n \dots A_{k+1}\| \leq c(\omega) + \xi_n + \dots + \xi_{k+1} + n\varepsilon$$

with probability 1.

Proof. We follow the proof of [37]. For a fixed $\varepsilon > 0$ take an l such that

$$\frac{1}{l} E \log \|A_l \dots A_1\| < \lambda + \varepsilon.$$

For any $0 \leq l < n$ and $0 \leq r < l$ take a cover of the index set $\{k+1, \dots, n\}$ by l -tuples of the form

$$I_q^r = \{ql + r, \dots, ql + r + l - 1\},$$

and let

$$\begin{aligned} q_0 &= q_0(r) = \min\{q : ql + r > k + 1\} \\ q_1 &= q_1(r) = \max\{q : ql + r \leq n\}. \end{aligned}$$

Set $\bar{k} = q_0l + r$ and $\bar{n} = q_1l + r - 1$. Then $\|A_n \dots A_{k+1}\|$ can be bounded from above by

$$\|A_n \dots A_{\bar{n}+1}\| \|A_{\bar{n}} \dots A_{\bar{n}-l+1}\| \dots \|A_{\bar{k}+l-1} \dots A_{\bar{k}}\| \|A_{\bar{k}-1} \dots A_{k+1}\|,$$

thus $\log \|A_n \dots A_{k+1}\|$ can be estimated from above by

$$\log \|A_n \dots A_{\bar{n}+1}\| + \sum_{q=q_0}^{q_1-1} \log \|A_{ql+r+l-1} \dots A_{ql+r}\| + \log \|A_{\bar{k}-1} \dots A_{k+1}\|. \quad (6.26)$$

By Lemma 6.2.1 we have that for any $\varepsilon' > 0$ there exists a random variable $c'(\omega)$ such that $\log \|A_k\| \leq c'(\omega) + \varepsilon'k$. Thus, since $n - (\bar{n} + 1) \leq l$ we have

$$\log \|A_n \dots A_{\bar{n}+1}\| \leq \log \|A_n\| + \dots + \log \|A_{\bar{n}+1}\| \leq l(c'(\omega) + \varepsilon'n).$$

A similar inequality holds for $\log \|A_{\bar{k}-1} \dots A_{k+1}\|$. Now define

$$\begin{aligned} \xi_i &= \log \|A_{i+l-1} \dots A_i\| \\ \eta_q^r &= \log \|A_{ql+r+l-1} \dots A_{ql+r}\|. \end{aligned}$$

Note that $\eta_q^r = \xi_{ql+r}$. Then the middle term in (6.26) can be written as

$$\sum_{q=q_0}^{q_1-1} \eta_q^r = \sum_{q=q_0}^{q_1-1} \xi_{ql+r}.$$

Letting r run from 0 to $(l-1)$ and averaging (6.26) over r we get

$$\log \|A_n \dots A_{k+1}\| \leq \left(\frac{1}{l} \sum_{i=k+1}^{n-l+1} \xi_i \right) + 2l(c'(\omega) + \varepsilon'n) \leq$$

$$\leq \left(\frac{1}{l} \sum_{i=k+1}^n \xi_i \right) + 3l(c'(\omega) + \varepsilon'(n+l)) \quad \text{a.s.},$$

since $\xi_i \leq l(c'(\omega) + \varepsilon'(n+l))$ for any $i \leq n$. For fixed l taking ε' sufficiently small we can upper bound the last, residual term as $c''(\omega) + \varepsilon n$ for any prescribed ε .

By assumption $\frac{1}{l}E\xi_i < \lambda + \varepsilon$, and (ξ_n) is strictly stationary and ergodic, hence, applying Lemma 6.2.4 we get that with some $c(\omega)$ depending on ε we have

$$\begin{aligned} \log \|A_n \dots A_{k+1}\| &\leq \frac{1}{l} \sum_{i=k+1}^n \xi_i + (c''(\omega) + \varepsilon n) \leq \\ &\leq c(\omega) + (n-k)(\lambda + \varepsilon) + n\varepsilon + (c''(\omega) + \varepsilon n) \quad \text{a.s.} \end{aligned}$$

which implies the claim. ■

From the above lemma it follows that, with $a_k = e^{\xi_k}$ we have

$$\|A_n \dots A_{k+1}\| \leq C(\omega) a_n \dots a_{k+1} e^{n\varepsilon},$$

and here $E \log a_n < \lambda(\mathcal{A}) + \varepsilon$.

Applying the above result for $\mathcal{A} = \mathcal{A}^2 = (A_n^2)$ in (6.22), the proof of Theorem 6.1.1 can be completed as in the scalar case.

6.2.2 Discussion on the proof of Theorem 6.1.1

In this subsection we present a few remarks to highlight the delicacy of the details of the proof of Theorem 6.1.1.

Remark 6.2.2 To estimate $\|A_n \dots A_{k+1}\|$, see Lemma 6.2.6, an alternative, direct approach would be to use the Fürstenberg-Kesten theorem starting at time k , and using the estimate that for all fixed $\varepsilon > 0$ we have

$$\|A_n \dots A_{k+1}\| \leq C_{k+1}(\omega) e^{(n-k)(\lambda + \varepsilon)},$$

where $\lambda = \lambda(\mathcal{A})$. Recall that $C_{k+1}(\omega)$ can be defined as $C_{k+1}(\omega) = e^{c_{k+1}(\omega)}$, where

$$c_{k+1}(\omega) = \sup_{n \geq k+1} (\log \|A_n \dots A_{k+1}\| - (n-k)(\lambda + \varepsilon)).$$

Using a representation of (A_n) via a measure-preserving shift on Ω it is easily seen that (C_{k+1}) can be assumed to be a stationary sequence. To control the effect of (C_{k+1}) in (6.22) we would need to show that (C_{k+1}) is sub-exponential. One way to show this would be to show that

$$\mathbb{E} \log^+ C_{k+1}(\omega) < +\infty. \quad (6.27)$$

Unfortunately, this inequality is not true in general. Indeed, consider a scalar valued i.i.d. process $A_n = a_n$. Then $\lambda = \mathbb{E} \log a_n$, and for fixed $\varepsilon > 0$ we have

$$\log C_{k+1}(\omega) = \sup_{n \geq k+1} \sum_{j=k+1}^n (\log a_j - (\lambda + \varepsilon)).$$

According to the Kiefer-Wolfowitz theorem (see Chow and Teicher [22], Chapter 10.4, Corollary 3) (6.27) holds if and only if

$$\mathbb{E}(\log^+ a_j)^2 < +\infty.$$

The same remark applies if we estimate $\|A_n \dots A_{k+1}\|$ using the Fürstenberg-Kesten theorem backwards in time. Then we get for fixed n , for all fixed $\varepsilon > 0$, and for all $k \leq n$

$$\|A_n \dots A_{k+1}\| \leq C_n(\omega) e^{(n-k)(\lambda + \varepsilon)},$$

and here the growth rate of $C_n(\omega)$ is, in general, not under control.

Remark 6.2.3 Note that in the case of *deterministic* matrices Theorem 6.1.1 does not hold. Indeed, defining a sequence of deterministic matrices in $\mathbb{R}^{2 \times 2}$ as

$$\bar{A}_n = \begin{pmatrix} A_n^1 & 0 \\ B_n & A_n^2 \end{pmatrix}$$

with $A_n^1 = 1 - \frac{\delta}{n}$, $0 < \delta < 1$ arbitrary, $A_n^2 = 1 - \frac{1}{n}$ and $B_n \equiv 1$, easy calculation yields that the right hand side of (6.22) cannot be estimated from above.

We note in passing, that an equally non-trivial deterministic version of Theorem 6.1.1, with A_n taking its values from a given set of matrices \mathcal{A} , has been given in Gerencsér and Michaletzky [69]. For more on this subject see Bara-

banov [4] and Dayawansa and Martin [27].

Remark 6.2.4 The existence of the derivative process $(X_{\theta,n})$ in an almost sure sense, and, as a byproduct, the existence of a strictly stationary causal solution of (6.7) has been proved by direct arguments in Berkes et al. [10] in the case of GARCH processes using specific arguments. Applying Theorem 3.2.2, we could conclude, with some additional work, that the top-Lyapunov exponent of the matrix

$$\bar{A}_n = \begin{pmatrix} A_n & 0 \\ A_{\theta,n} & A_n \end{pmatrix}.$$

is negative. It is not clear if this line of proof can be used in the general case.

6.2.3 Proof of Theorem 6.1.2

First we consider the case $q = 2$. It is easy to see that $A \otimes A$ is block-triangular with blocks in the diagonal $A_1 \otimes A$ and $A_2 \otimes A$, and with the block $B \otimes A$ in position $(2, 1)$.

Let the zero block of A in the position $(1, 2)$ be denoted by $A_{1,2}$. To compute $\det(A_i \otimes A - \lambda I)$, with $i = 1, 2$, we permute the columns and rows of $A_i \otimes A$ so that the zero sub-matrices of the form $a_{i,rs}A_{1,2}$ are merged into a single zero block of $A_i \otimes A$ in the $(1, 2)$ position. We describe the procedure for $i = 2$. Let us partition $A_2 \otimes A$ according to the partition of A into blocks of rows and columns of alternating widths d_1 and d_2 . The number of blocks of rows and block columns thus will be $2d_2$. Moving a block row of index $2i$, $i \leq d_2$ to position $d_2 + i$, and proceeding with columns similarly it is easy to see that this permutation of block rows and block columns of $A_2 \otimes A - \lambda I$ yields the matrix

$$\begin{pmatrix} A_2 \otimes A_1 - \lambda I & 0 \\ A_2 \otimes B & A_2 \otimes A_2 - \lambda I \end{pmatrix}.$$

Remark 6.2.5 We can summarize the above procedure as follows: if $A^{(1)} = A^{(2)} = A$ are block-triangular matrices, then there exist permutation matrix $\Pi^{(2)}$ acting on the rows and columns of $A^{(2)}$ such that

$$A^{(1)} \otimes (\Pi^{(2)} A^{(2)} \Pi^{(2)}) = (I \otimes \Pi^{(2)}) \cdot (A^{(1)} \otimes A^{(2)}) \cdot (I \otimes \Pi^{(2)}) = \Pi \cdot (A^{(1)} \otimes A^{(2)}) \cdot \Pi$$

is block diagonal with blocks of the form $(A_{i_1}^{(1)} \otimes A_{i_2}^{(2)})$, with $i_1, i_2 = 1, 2$. Obviously the diagonal of $A^{(1)} \otimes A^{(2)}$ will be mapped into the diagonal of $\Pi \cdot (A^{(1)} \otimes A^{(2)}) \cdot \Pi$. Consequently,

$$\Pi \cdot (A^{(1)} \otimes A^{(2)} - \lambda I) \cdot \Pi,$$

with I denoting a unit matrix of appropriate dimension, will be block diagonal with blocks of the form

$$(A_{i_1}^{(1)} \otimes A_{i_2}^{(2)}) - \lambda I,$$

with $i_1, i_2 = 1, 2$.

Now permutation and taking expectation commute. Thus the determinant of $E(A_2 \otimes A - \lambda I)$ will be the same as the determinant of the expectation of the matrix displayed above. Letting $\text{Sp}(A)$ denote the spectrum of the matrix A with multiplicity taken into account we conclude that

$$\text{Sp} [E(A_2 \otimes A)] = \text{Sp} [E(A_2 \otimes A_1)] \cup \text{Sp} [E(A_2 \otimes A_2)] .$$

The $(1, 1)$ block of $A \otimes A$ is handled similarly using $A^T \otimes A^T$, hence we get that

$$\text{Sp} [E(A_1 \otimes A)] = \text{Sp} [E(A_1 \otimes A_1)] \cup \text{Sp} [E(A_1 \otimes A_2)] .$$

Thus it is sufficient to show that

$$\rho [E(A_2 \otimes A_1)] \leq \max\{\rho [E(A_1 \otimes A_1)] ; \rho [E(A_2 \otimes A_2)] \}.$$

This follows from the following general result.

Lemma 6.2.7 *Let U, V be random square matrices of arbitrary dimensions. Then*

$$\{\rho [E(U \otimes V)] \}^2 \leq \rho [E(U \otimes U)] \cdot \rho [E(V \otimes V)] .$$

The above inequality is a Cauchy-Schwartz inequality stated for tensor-products. For its proof we need another variant of the Cauchy-Schwartz inequality stated for tensor-products.

Lemma 6.2.8 *Let U, V be random square matrices of arbitrary dimensions. Then*

$$\|E(U \otimes V)\|^2 \leq \|E(U \otimes U)\| \cdot \|E(V \otimes V)\|,$$

where $\|\cdot\|$ denotes the operator norm.

Proof of Lemma 6.2.7: Recall that for any square matrix C we have

$$\rho(C) = \lim_{m \rightarrow \infty} \|C^m\|^{1/m}.$$

Take $C = E(U \otimes V)$, and let (U_i, V_i) be m independent copies of (U, V) . Then

$$\begin{aligned} C^m &= [E(U \otimes V)]^m = E[(U_m \otimes V_m) \dots (U_1 \otimes V_1)] = \\ &= E[U_m \dots U_1 \otimes V_m \dots V_1]. \end{aligned} \quad (6.28)$$

Thus, by Lemma 6.2.8,

$$\begin{aligned} \|C^m\|^2 &= \|E[U_m \dots U_1 \otimes V_m \dots V_1]\|^2 \leq \\ &\leq \|E[U_m \dots U_1 \otimes U_m \dots U_1]\| \cdot \|E[V_m \dots V_1 \otimes V_m \dots V_1]\|. \end{aligned} \quad (6.29)$$

But

$$E[U_m \dots U_1 \otimes U_m \dots U_1] = E(U_m \otimes U_m) \dots (U_1 \otimes U_1) = [E(U \otimes U)]^m,$$

and similarly for the second term in the right hand side of (6.29). Taking m -th root of both sides, and taking limit for $m \rightarrow \infty$ we get the claim. \blacksquare

Proof of Lemma 6.2.8: Recall that for any matrix C we have

$$\|C\|^2 = \max_{|x|=1} |Cx|^2 = \max_{|x|=1} x^T C^T C x = \|C^T C\|.$$

Take $C = E(U \otimes V)$, and let (\bar{U}, \bar{V}) be an independent copy of (U, V) . Then

$$\begin{aligned} C^T C &= E(\bar{U} \otimes \bar{V})^T \cdot E(U \otimes V) \\ &= E[(\bar{U}^T \otimes \bar{V}^T)(U \otimes V)] = E(\bar{U}^T U \otimes \bar{V}^T V). \end{aligned} \quad (6.30)$$

In the same way

$$\mathbf{E}(\bar{U} \otimes \bar{U})^T \cdot \mathbf{E}(U \otimes U) = \mathbf{E}(\bar{U}^T U \otimes \bar{U}^T U)$$

and similarly for V .

Obviously, Lemma 6.2.8 for given (U, V) is equivalent to Lemma 6.2.8 for (U, V) being replaced by $(\bar{U}^T U, \bar{V}^T V)$. Thus it is sufficient to handle the latter case. From now on we assume that U, V themselves are square matrices, and that $\mathbf{E}(U \otimes V)$, $\mathbf{E}(U \otimes U)$ and $\mathbf{E}(V \otimes V)$ are all symmetric. We can also assume that U, V have the same dimension: if this is not the case we augment the smaller matrix with zeros. To handle this case we need one more extension of the Cauchy-Schwartz inequality for tensor-products.

Lemma 6.2.9 *Let U, V be random square matrices of identical dimensions, say n . Then*

$$[\operatorname{tr} \mathbf{E}(U \otimes V)]^2 \leq \operatorname{tr} \mathbf{E}(U \otimes U) \cdot \operatorname{tr} \mathbf{E}(V \otimes V).$$

Proof of Lemma 6.2.9: The function

$$f(U, V) = \operatorname{tr} \mathbf{E}(U \otimes V) = \mathbf{E} \operatorname{tr}(U \otimes V) = \mathbf{E}(\operatorname{tr} U \cdot \operatorname{tr} V)$$

is symmetric, positive semi-definite function on the linear space of $n \times n$ random matrices. Hence Lemma 6.2.9 is nothing else, but the Cauchy-Schwartz inequality for this inner product space. ■

Let U, V be again square matrices of identical dimensions, such that $\mathbf{E}(U \otimes V)$, $\mathbf{E}(U \otimes U)$ and $\mathbf{E}(V \otimes V)$ are all symmetric. Let

$$\operatorname{Sp} \mathbf{E}(U \otimes U) = (\lambda_i)$$

$$\operatorname{Sp} \mathbf{E}(V \otimes V) = (\mu_i)$$

$$\operatorname{Sp} \mathbf{E}(U \otimes V) = (\rho_i).$$

Due to the assumed symmetry, all eigenvalues are real. It can be also assumed that all of the eigenvalues mentioned above are non-negatives, since the the positive semi-definite property and the symmetry of the matrices $U \otimes U, V \otimes V$

and $U \otimes V$ implies that $E(U \otimes V)$, $E(U \otimes U)$ and $E(V \otimes V)$ are all positive semi-definite matrices. Then Lemma 6.2.9 can be written as

$$\left(\sum_i \rho_i\right)^2 \leq \left(\sum_i \lambda_i\right)\left(\sum_i \mu_i\right).$$

Lemma 6.2.10 *Let m be a positive integer. Then*

$$\left(\sum_i \rho_i^m\right)^2 \leq \left(\sum_i \lambda_i^m\right)\left(\sum_i \mu_i^m\right).$$

Proof of Lemma 6.2.10: Let $C = E(U \otimes V)$. Then

$$\sum_i \rho_i^m = \text{tr } C^m.$$

But, following the arguments given to derive (6.28),

$$C^m = [E(U \otimes V)]^m = E[U_m \dots U_1 \otimes V_m \dots V_1].$$

Thus

$$\text{tr } C^m = \text{tr } E[U_m \dots U_1 \otimes V_m \dots V_1].$$

We can express $\sum_i \lambda_i^m$ and $\sum_i \mu_i^m$ similarly. Applying Lemma 6.2.9 with U, V being replaced by $U_m \dots U_1$ and $V_m \dots V_1$, respectively, gives the claim. ■

Proof of Lemma 6.2.8 (continued): Let

$$\|E(U \otimes U)\| = \max_i |\lambda_i| = \lambda^*$$

$$\|E(V \otimes V)\| = \max_i |\mu_i| = \mu^*$$

$$\|E(U \otimes V)\| = \max_i |\rho_i| = \rho^*.$$

Taking the m -th root of both sides in Lemma 6.2.10, and taking the limit for $m \rightarrow \infty$ which exists due to the non-negativity assumption on the eigenvalues, we get

$$(\rho^*)^2 \leq \mu^* \lambda^*,$$

which proves the claim for $q = 2$. ■

In the general case, we have for any $q \geq 2$, by a simple extension of the arguments of Remark 6.2.5, that

$$\text{Sp} [E(A^{\otimes q})] = \cup_{i_q, \dots, i_1} \text{Sp} [E(A_{i_q} \otimes \dots \otimes A_{i_1})],$$

with multiplicity taken into account, where $i_s = 1, 2$ for $s = 1, \dots, q$. Thus it is sufficient to show that for any (i_q, \dots, i_1) we have

$$\rho [E(A_{i_q} \otimes \dots \otimes A_{i_1})] \leq \max\{ \rho [E(A_1^{\otimes q})] ; \rho [E(A_2^{\otimes q})] \}. \quad (6.31)$$

Setting

$$C = C_{i_q, \dots, i_1} = E(A_{i_q} \otimes \dots \otimes A_{i_1})$$

the estimation of $\rho(C)$ is reduced to the estimation of

$$\text{tr}[C^m (C^T)^n]^r = \text{tr}[C^m (C^T)^n \cdot \dots \cdot C^m (C^T)^n]$$

as in the case $q = 2$. Now take $2nr$ independent copies of $(A_{i_q}, \dots, A_{i_1})$, say $(A_{i_q}, \dots, A_{i_1})^{(t)}$, with $t = 1, \dots, 2nr$. Then we can write

$$[C^m (C^T)^n]^r = E(F_{i_q} \otimes \dots \otimes F_{i_1}),$$

where for any $s = 1, \dots, q$ we have

$$F_{i_s} = A_{i_s}^{(2nr)} \cdot \dots \cdot A_{i_s}^{(2nr-n+1)} \cdot \dots \cdot (A_{i_s}^T)^{(n)} \cdot \dots \cdot (A_{i_s}^T)^{(1)}.$$

Now

$$\text{tr} E(F_{i_q} \otimes \dots \otimes F_{i_1}) = E \text{tr}(F_{i_q} \otimes \dots \otimes F_{i_1}) = E \Pi_{s=1}^q \text{tr}(F_{i_s}).$$

Applying Hölder's inequality with exponents q for each term, we get the upper bound

$$\Pi_{s=1}^q E^{1/q} [\text{tr}(F_{i_s})]^q \leq \max\{ E [\text{tr}(F_1)]^q, E [\text{tr}(F_2)]^q \}.$$

Noting that for $i = 1, 2$

$$\mathbb{E} [\mathrm{tr}(F_i)]^q = \mathrm{tr} \, \mathbb{E}(F_i \otimes \dots \otimes F_i) = C_{i, \dots, i} = \mathbb{E}(A_i \otimes \dots \otimes A_i) \, ,$$

the proposition follows. ■

Chapter 7

Analysis of the recursive algorithm

The main contribution of this chapter is a rigorous convergence analysis of the recursive estimation method for the parameters of GARCH processes, proposed in Chapter 4, with large stability margin, under reasonable technical conditions.

The first section of this chapter is devoted to our main result on the convergence of the proposed recursive algorithm for GARCH processes, while the proof is given in Section 7.2. The analysis is equally applicable to (4.12)-(4.13). All of the results of this chapter are based on the article Gerencsér and Orlovits [46] and the conference proceeding Gerencsér et al. [47]. Our results complement the results of Aknouche and Guerbyenne [1] and Dahlhaus and Subba Rao [25] in the sense that our algorithm will converge in almost sure sense as well as in L_q up to certain q -s. The viability of the method will be demonstrated in Section 7.3 by experimental results both for simulated and real data.

7.1 The algorithm and the main result

Recall that for the solution of the general estimation problem (4.4) defined in Chapter 4 the following stochastic approximation procedure is proposed: starting with some initial condition $\theta_0 \in K_0$ we define recursively

$$\theta_n = \theta_{n-1} - \frac{1}{n} \frac{\sigma_{\theta,n}}{\sigma_n} \left(1 - \frac{y_n^2}{\sigma_n^2} \right), \quad (7.1)$$

where σ_n and $\sigma_{\theta,n}$ denote the on-line estimates of $\bar{\sigma}_n(\theta_{n-1})$ and $\bar{\sigma}_{\theta,n}(\theta_{n-1})$, respectively. Recall also from Chapter 4, that the problem of estimating the GARCH parameters has been formulated into a form of a linear stochastic system

$$\bar{\psi}_{n+1}(\theta) = P_{n+1}(\theta)\bar{\psi}_n(\theta) + w_n(\theta) \quad (7.2)$$

with a *block-triangular* state transition matrix

$$P_n(\theta) = \begin{pmatrix} A_n^e(\theta) & 0 \\ x & \tilde{A}_n^e(\theta) \end{pmatrix},$$

It is obvious that $\bar{\psi}_n(\theta)$ is (a parameter dependent) Markov process. Since the asymptotic estimation problem (4.7) can be formulated in terms of $\bar{\psi}_n(\theta)$, the BMP-theory is applicable, and we end up with the SA procedure (7.1). To prove the convergence of (7.1), suitably modified with resetting, we need to verify the general conditions of the BMP-theory given above.

To simplify the notations we drop the dependence on θ . The technically most demanding condition of the BMP-theory is Condition 5.2.2. For its verification we prove that the q -th mean Lyapunov exponent associated with the i.i.d. sequence of matrices $\mathcal{P} = (P_n)$, and defined in (3.17) as

$$\lambda_q(\mathcal{P}) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} \|P_n \dots P_1\|^q,$$

is strictly negative. A sufficient condition for the negativity of $\lambda_q(\mathcal{P})$ is formulated in Theorem 3.3.1 of Chapter 3, namely

$$\rho[\mathbb{E}(P_0)^{\otimes q}] < 1 \quad (7.3)$$

implies that $\lambda_q(\mathcal{P}) < 0$. The verification of (7.3) can be significantly simplified by exploiting the *block-triangular* structure of P_n and applying Theorem 6.1.2. This yields the following corollary:

Corrolary 7.1.1 *Let q be an integer, even or odd, and let $D(z^{-1})$ be stable. Then*

$$\rho[\mathbb{E}(A_0^*)^{\otimes q}] < 1 \quad \text{implies} \quad \rho[\mathbb{E}(P_0)^{\otimes q}] < 1.$$

If q is even, then it follows that

$$\lambda_q = \lambda_q(\mathcal{P}) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} \|P_n \dots P_1\|^q < 0.$$

It follows that for any $\varepsilon > 0$ we have

$$\mathbb{E} \|P_n \dots P_1\|^q \leq C e^{(\lambda_q + \varepsilon)n}$$

with some $C = C(\varepsilon) > 0$.

We are now ready to state our main theorem:

Theorem 7.1.1 *Let $D^*(z^{-1})$ be stable, and let ε , the limiting rate of θ_n , be sufficiently small. Let the truncation domain D_0 be such that for all $\theta \in D_0$ the corresponding polynomial $D(z^{-1})$ is stable, and let Condition 5.4.1 be satisfied. Assume further that with some even $q \geq 6$*

$$\rho[\mathbb{E}(A_0^*)^{\otimes q}] < 1 \tag{7.4}$$

is satisfied. Then the estimator sequence θ_n given by (7.1), and modified by a resetting mechanism converges to θ^ w.p.1, and also in L_q , with rate*

$$\mathbb{E}^{1/q} |\theta_n - \theta^*|^q = O(n^{-(\alpha \wedge \frac{1}{2})}),$$

where $-\alpha < 0$ is the Lyapunov-exponent of the associated ODE.

Remark 7.1.1 Recall from Remark 5.4.1 that the Lyapunov-exponent of the associated ODE is given by the maximum of the real-parts of the eigenvalues of the Jacobian-matrix of the right hand side at $\theta = \theta^*$. An analogous theorem is valid for the stochastic Newton method, in which case $\alpha = -1$, and the convergence rate in L_q is $O(n^{-\frac{1}{2}})$.

Remark 7.1.2 The verification of (7.4) is not easy. A rough upper bound for $\rho[\mathbb{E}(A_0^*)^{\otimes q}]$ is given by the following inequality: assuming

$$\sum_{i=1}^r \alpha_i^* + \sum_{j=1}^s \beta_j^* < 1,$$

and setting $A^* := E(A_0^*)$, we have, for any even q ,

$$\rho[E(A_0^*)^{\otimes q}] \leq \rho[E(\bar{\varepsilon}_0^2 \cdot I)^{\otimes q} \cdot (A^*)^{\otimes q}] \leq E(\bar{\varepsilon}_0^{2q})\rho(A^*)^q. \quad (7.5)$$

Here we used the special structure of the A_n^* , and the properties of the Kronecker product. Thus $\rho[E(A_0^*)^{\otimes q}] < 1$ is satisfied if

$$\rho(A^*) < \frac{1}{E^{1/q}(\bar{\varepsilon}_0^{2q})}. \quad (7.6)$$

In spite of the difficulty of verifying this condition simulation results show that our algorithm performs very well even for problems with stability margin as small as 1%, see Section 7.3.

Remark 7.1.3 Note that, by the Perron-Frobenius theorem, the stability condition

$$\sum_{i=1}^r \alpha_i^* + \sum_{j=1}^s \beta_j^* < 1,$$

implies that $\rho(A^*) < 1$. Since $E^{1/q}(\bar{\varepsilon}_0^{2q}) > 1$, for $q \geq 1$, condition (7.6) requires that A^* has a restricted stability margin. This may put a severe limitation on the range of applicability of Theorem 7.1.1. The extent of this restriction will be discussed below. For the verification of (7.6) a good upper bound for $\rho(A^*)$ is given in Theorem 1 of Stefanescu [83]. Note that the recursive algorithm proposed in Aknouché and Guerbyenne [1] has a limited range of applicability, as well, namely it is required that the following positive real condition is satisfied, with D defined in (3.3):

$$\operatorname{Re}[D(e^{i\omega})^{-1} - 1/2] > 0. \quad (7.7)$$

Remark 7.1.4 The asymptotic covariance matrix of recursive estimator is known to satisfy a certain Lyapunov-equation, see Benveniste et al. [7], under appropriate technical conditions. It is then a well-known fact that for a Newtonian recursive estimator the asymptotic covariance matrix of recursive estimator is the same as the asymptotic covariance matrix of the off-line esti-

mator. The latter is given above. It is likely that these results of Benveniste et al. [7], can be extended to our modified algorithm which uses resetting.

7.2 Verification of conditions for convergence

In this section we verify the conditions of the general BMP theory developed in Chapter 5 for the convergence of the proposed algorithm. The hard part is to verify the conditions for the Markov kernel. First note that the conditions of Theorem 7.1.1 imply that $\varepsilon_n \in L_q$, hence, for the state-transition matrix and the input in (7.2) we trivially have $P_1(\theta), w_1(\theta) \in L_q$. The validity of Condition 5.2.2, ensuring among others that the state process $\bar{\psi}_n(\theta)$ is L_q -bounded, is stated in the following lemma:

Lemma 7.2.1 *Assume that the conditions of Theorem 7.1.1 are satisfied. Then Condition 5.2.2 holds with $\bar{X}_n(\theta) = \bar{\psi}_n(\theta)$.*

Proof. Using the state space form (7.2), we have that

$$\bar{\psi}_1(\theta) = P_1(\theta)\bar{\psi}_0(\theta) + w_1(\theta).$$

Thus, taking L_q -norm on both sides and applying the triangle inequality yields

$$\mathbb{E}^{1/q}(1 + |\bar{\psi}_1(\theta)|^q) \leq 1 + \mathbb{E}^{1/q}|P_1(\theta)|^q|\psi|^q + \mathbb{E}^{1/q}|w_1(\theta)|^q \leq K(1 + |\psi|^q)$$

with $\psi = \bar{\psi}_0(\theta)$ being a non-random arbitrary initial value of the process $\bar{\psi}_n(\theta)$. Thus the first part of Condition 5.2.2 is verified. For the second part, express $\bar{\psi}_r(\theta)$ by iterating the extended state equation (7.2) to get:

$$\bar{\psi}_r(\theta) = (P_r(\theta) \dots P_1(\theta))\bar{\psi}_0(\theta) + \sum_{k=0}^{r-1} (P_r(\theta) \dots P_{r-k+1}(\theta))w_{r-k}(\theta). \quad (7.8)$$

Applying the triangle inequality on the right-hand side of (7.8) we get that $\mathbb{E}^{1/q}|\bar{\psi}_r(\theta)|^q$ can be estimated from above by

$$\mathbb{E}^{1/q}\|P_r(\theta) \dots P_1(\theta)\|^q|\psi|^q + \sum_{k=0}^{r-1} \mathbb{E}^{1/q}|P_r(\theta) \dots P_{r-k+1}(\theta)w_{r-k}(\theta)|^q.$$

For the k -th term on the right hand side, we apply the inequality $|Ax| \leq \|A\||x|$, and noting that $\|P_r(\theta) \dots P_{r-k+1}(\theta)\|^q$ and $|w_{r-k}(\theta)|^q$ are *independent*, we get the upper bound

$$\mathbb{E}^{1/q} \|P_r(\theta) \dots P_{r-k+1}(\theta)\|^q \mathbb{E}^{1/q} |w_{r-k}(\theta)|^q. \quad (7.9)$$

Now, we have $\mathbb{E}^{1/q} |w_{r-k}(\theta)|^q \leq K_1 < +\infty$. Applying Corollary 7.1.1, and taking into account that the $P_i(\theta)$ -s are i.i.d., we get that for any $\varepsilon > 0$

$$\mathbb{E}^{1/q} \|P_r(\theta) \dots P_{r-k+1}(\theta)\|^q \leq C e^{(\lambda_q(\mathcal{P})+\varepsilon)r/q} := C \rho^r \quad (7.10)$$

with some finite constant $C = C(\varepsilon) > 0$, $0 < \rho < 1$ and $\lambda_q(\mathcal{P}) < 0$. Thus we get that

$$\mathbb{E}^{1/q} |\bar{\psi}_r(\theta)|^q \leq C \rho^r |\psi|^q + \sum_{k=0}^{r-1} C' \rho^k K_1,$$

from which, with sufficiently large r , the claim follows. \blacksquare

Lemma 7.2.2 below addresses a certain forgetting property of the Markov transition kernel Π_θ formulated in general in Condition 5.2.1. Let $\bar{\psi}_n(\theta)$ and $\bar{\psi}'_n(\theta)$ denote the frozen-parameter processes with initial values $\bar{\psi}_0(\theta) = \psi$ and $\bar{\psi}'_0(\theta) = \psi'$, respectively.

Lemma 7.2.2 *Assume that the conditions of Theorem 7.1.1 are satisfied. Then Condition 5.2.1 holds with $\bar{X}_n(\theta) = \bar{\psi}_n(\theta)$.*

Proof. Since

$$|\Pi_\theta^n g(\psi) - \Pi_\theta^n g(\psi')| = |\mathbb{E}[g(\bar{\psi}_n(\theta)) - g(\bar{\psi}'_n(\theta))]| \quad (7.11)$$

and $g \in Li(q)$, we get that the right hand side of (7.11) can be estimated from above by

$$\|\Delta g\|_q \mathbb{E}[|\bar{\psi}_n(\theta) - \bar{\psi}'_n(\theta)|(1 + |\bar{\psi}_n(\theta)|^q + |\bar{\psi}'_n(\theta)|^q)]. \quad (7.12)$$

From the state equation (7.2) it follows that

$$|\bar{\psi}_n(\theta) - \bar{\psi}'_n(\theta)| \leq \|P_n(\theta) \dots P_1(\theta)\| |\psi - \psi'|,$$

thus for the second part of (7.12) we get the upper bound

$$|\psi - \psi'|E\left[\|P_n(\theta) \dots P_1(\theta)\|(1 + |\bar{\psi}_n(\theta)|^q + |\bar{\psi}'_n(\theta)|^q)\right]. \quad (7.13)$$

Applying the Hölder inequality with the conjugate exponents $l = q + 1, m = (q + 1)/q$ yields that the second term of (7.13) can be estimated from above by

$$E^{1/l}\|P_n(\theta) \dots P_1(\theta)\|^l E^{1/m}(1 + |\bar{\psi}_n(\theta)|^q + |\bar{\psi}'_n(\theta)|^q)^m. \quad (7.14)$$

Estimating the first term of (7.14) we can use also Corollary 7.1.1. For the last term of (7.14) we apply the triangle inequality and the second statement of Lemma 7.2.1 which yields

$$E^{1/m}(1 + |\bar{\psi}_n(\theta)|^q + |\bar{\psi}'_n(\theta)|^q)^m \leq C(1 + |y|^q + |y'|^q) \quad (7.15)$$

with some finite constant C . Taking into account the above estimations yields the statement. \blacksquare

The next statement will provide that the kernels Π_θ^n are Lipschitz continuous, uniformly in n , with respect to the parameter θ when applied to the set of test functions $Li(q)$. Now let $\bar{\psi}_n(\theta_1)$ and $\bar{\psi}_n(\theta_2)$ denote the frozen parameter processes starting with the same non-random initial state $\bar{\psi}_0(\theta_1) = \bar{\psi}_0(\theta_2) = \psi$.

Lemma 7.2.3 *Assume that the conditions of Theorem 7.1.1 are satisfied. Then Condition 5.2.3 holds with $\bar{X}_n(\theta) = \bar{\psi}_n(\theta)$.*

Proof. Since

$$|\Pi_{\theta_1}^n g(\psi) - \Pi_{\theta_2}^n g(\psi)| = |E[g(\bar{\psi}_n(\theta_1)) - g(\bar{\psi}_n(\theta_2))]| \quad (7.16)$$

and $g \in Li(q)$, we get that (7.16) is majorated by

$$\|\Delta g\|_q E\left[|\bar{\psi}_n(\theta_1) - \bar{\psi}_n(\theta_2)|(1 + |\bar{\psi}_n(\theta_1)|^q + |\bar{\psi}_n(\theta_2)|^q)\right]. \quad (7.17)$$

Applying the Hölder inequality with the conjugate exponents $l = q + 1, m = (q + 1)/q$ yields that the second term of (7.17) can be estimated from above by

$$E^{1/l} |\bar{\psi}_n(\theta_1) - \bar{\psi}_n(\theta_2)|^l E^{1/m} (1 + |\bar{\psi}_n(\theta_1)|^q + |\bar{\psi}_n(\theta_2)|^q)^m. \quad (7.18)$$

Using the iterated state equation (7.8) for both $\bar{\psi}_n(\theta_1)$ and $\bar{\psi}_n(\theta_2)$ and applying the triangle inequality yields that $E^{1/l} |\bar{\psi}_n(\theta_1) - \bar{\psi}_n(\theta_2)|^l$ can be bounded from above by

$$E^{1/l} \left\| \prod_{k=1}^n P_k(\theta_1) - \prod_{k=1}^n P_k(\theta_2) \right\|^l \cdot |\psi| + E^{1/l} |U_n(\theta_1) - U_n(\theta_2)|^l, \quad (7.19)$$

with

$$U_n(\theta_i) = \sum_{k=0}^{n-1} P_n(\theta_i) \dots P_{n-k+1}(\theta_i) w_{n-k}(\theta_i), \quad i = 1, 2.$$

Lemma 7.2.4 *Under the conditions of Lemma 7.2.3 there exist constants C such that*

$$E^{1/l} |\bar{\psi}_n(\theta_1) - \bar{\psi}_n(\theta_2)|^l \leq C |\theta_1 - \theta_2| (1 + |\psi|) \quad (7.20)$$

The proof of this statement is given in Appendix B. Here we only mention that its proof is based on the following useful lemma. Its proof is also relegated into Appendix B.

Lemma 7.2.5 *Under the conditions of Lemma 7.2.3 there exist constants C, K and $0 < \rho < 1$ such that*

$$E^{1/l} \left\| \prod_{k=1}^n P_k(\theta_1) - \prod_{k=1}^n P_k(\theta_2) \right\|^l \leq K \rho^n |\theta_1 - \theta_2| \quad (7.21)$$

To estimate the last term of (7.18) we can use also the triangle inequality and the second statement of Lemma 7.2.1 with $m = (q + 1)/q$, from which the claim follows. ■

With this all conditions of the BMP theory on the Markov kernel has been verified.

The correction term in our SA algorithm (7.1) is defined via the function

$$H(\theta; y^2, \sigma, \sigma_\theta) = \frac{\sigma_\theta}{\sigma} \left[1 - \frac{y^2}{\sigma^2} \right]. \quad (7.22)$$

Lemma 7.2.6 *Consider the GARCH(r, s) process defined by equations (3.1) and (3.2) with $\alpha_0 > 0$ and $\alpha_i, \beta_j \geq 0$, $i = 1, \dots, r$, $j = 1, \dots, s$. Then Condition 5.2.4 holds with $p = 1$.*

Proof. Since H is quadratic in the variable $z = (y^2, \sigma, \sigma_\theta)$ and $\sigma^2 \geq \alpha_0 > 0$, it is straightforward to see that the first and second part of Condition 5.2.4 are trivially fulfilled with $p = 1$. The third part of Condition 5.2.4 follows immediately with $p = 1$ from the application of the mean value theorem. ■

Proof of Theorem 7.1.1: Lemma 7.2.3 and the condition $q > 2(p+1) = 4$, q even, in Condition 5.2.4 brings us to the assumption $q \geq 6$. Applying Lemma 7.2.1 with this q ensures that Condition 5.2.2 is satisfied for $\bar{X}_n(\theta) = \bar{\psi}_n(\theta)$ with $q = 6$. Using Condition 5.2.2 with $q = 6$ and Lemma 7.2.2 and 7.2.3, the validity of Condition 5.2.1 and 5.2.3 for $\bar{X}_n(\theta) = \bar{\psi}_n(\theta)$ follows immediately. Hence we get that Theorem 5.4.1 is applicable for $\bar{X}_n(\theta) = \bar{\psi}_n(\theta)$ with $q \geq 6$, which proves the theorem.

7.3 Simulation results

In this section we test the performance of our algorithm via simulated data. For all simulations we use a generated data set with ten thousand observations. The figure below shows the trajectory of the estimated parameter process θ_n for 10,000 simulated observations generated by a GARCH(1, 1) model driven by Gaussian white noise, with zero mean and unit variance, and with parameters

$$\theta^* = \begin{pmatrix} 0.3 \\ 0.05 \\ 0.8 \end{pmatrix}. \quad (7.23)$$

The stability margin of this system is 15%. The recursive estimators are depicted on the figure below.

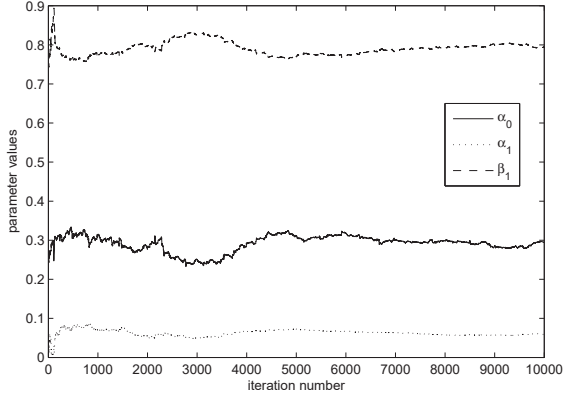


Figure 7.1: Trajectory of $\hat{\theta}_n$ vs. n with initial values (0.25, 0.01, 0.75)

The inverse of the estimated Fisher information matrix based on an empirical version of (4.10), using on line estimates of $\bar{\sigma}_n(\theta^*)$ and $\bar{\sigma}_{\theta,n}(\theta^*)$, is

$$\hat{I}^{-1} = \begin{pmatrix} 24.6682 & 1.8385 & -14.2332 \\ 1.8385 & 0.6601 & -1.5368 \\ -14.2332 & -1.5368 & 8.7286 \end{pmatrix}.$$

Similar findings were found in the paper of Gerencsér, Orlovits and Torma [47]. It is easy to see that the Fisher information matrix is ill-conditioned, which can be visually explained by the presence of a ridge in the asymptotic cost function.

In comparing our results with the classical off-line conditional maximum-likelihood (ML) method we recompute the inverse of the estimated Fisher information matrix based on an empirical version of (4.10), using exact data $\bar{\sigma}_n(\theta^*)$ and $\bar{\sigma}_{\theta,n}(\theta^*)$. The next table presents the final estimates after 10,000 samples results for both the off-line and the on-line estimation method. The estimated standard errors of the estimated parameters are given in parenthesis.

Table I. Comparison of off-line and on-line estimates

	ML off-line	on-line
$\hat{\alpha}_0$	0.3501 (0.0691)	0.3004 (0.0424)
$\hat{\alpha}_1$	0.0643 (0.0398)	0.0580 (0.0073)
$\hat{\beta}_1$	0.7639 (0.0096)	0.7931 (0.0244)

Explanation of the Tables. First, the sensitivity with respect to the initial values of the parameters is analyzed for the above model (7.23). After a large number of random trials, four possible initial values θ_0 of the parameters were chosen within the stability region, as shown in first line of Table II. In each column the estimated final values of the parameters are given, after 10,000 iterations. The estimated standard errors of the estimated parameters are given in parenthesis.

Table II. Estimated parameters of the GARCH(1,1) model (7.23).

θ_0	0.33 0.055 0.88	0.27 0.045 0.72	0.21 0.035 0.56	0.15 0.025 0.4
$\hat{\alpha}_0$	0.3515 (0.0442)	0.3172 (0.0747)	0.2877 (0.0420)	0.2534 (0.0274)
$\hat{\alpha}_1$	0.0403 (0.0068)	0.0485 (0.0081)	0.0468 (0.0075)	0.0535 (0.0061)
$\hat{\beta}_1$	0.8381 (0.0252)	0.7917 (0.0422)	0.8106 (0.0246)	0.8413 (0.0175)

These empirical studies indicate that the accuracy of the final estimator depends strongly on the initial values of the estimator. This may be partially due to the non-convexity and flatness of the asymptotic cost function. Since

we can assume that recursive estimation is primarily used for keeping track of eventual small changes in a priori more or less known parameters, in the next two examples we consider initial values close to the true values.

In the next two examples we consider GARCH processes with stability margins as small as 2% and 1%, which are consistent with our findings for *real data*, see below. In both examples we chose the initial values by simultaneously moving up or down the true parameters in various combinations. In the up-moves we take a fraction of the stability margin, in the down-moves we allow approximately 10% perturbation. A model with stability margin $1 - \alpha_1^* - \beta_1^* = 0.02$ is given by the following parameters:

$$\theta^* = \begin{pmatrix} 0.001 \\ 0.08 \\ 0.9 \end{pmatrix}. \quad (7.24)$$

The results are summarized in the following table following the structure given in the explanations preceding Table II:

Table III. Estimated parameters of the GARCH(1, 1) model (7.24).

	0.0002	0.0006	0.0014	0.0018
θ_0	0.0720	0.0680	0.0849	0.0898
	0.7200	0.9259	0.9049	0.9098
$\hat{\alpha}_0$	0.0009	0.0008	0.0009	0.0010
	(0.0001)	(0.0001)	(0.0001)	(0.0002)
$\hat{\alpha}_1$	0.0815	0.0784	0.0827	0.0820
	(0.0018)	(0.0044)	(0.0053)	(0.0051)
$\hat{\beta}_1$	0.9081	0.9045	0.9007	0.9017
	(0.0019)	(0.0054)	(0.0114)	(0.0258)

Our second model with stability margin $1 - \alpha_1^* - \beta_1^* = 0.01$ is given by:

$$\theta^* = \begin{pmatrix} 0.001 \\ 0.09 \\ 0.9 \end{pmatrix}. \quad (7.25)$$

The results are summarized in the following table. Note that in both cases the proposed algorithm performs remarkably well in spite of the fact that a condition of Theorem 4, notably (7.4) is unlikely to be satisfied even with $q = 6$, since $E\varepsilon_0^6 = 15$.

Table IV. Estimated parameters of the GARCH(1, 1) model (7.25).

θ_0	0.0002	0.0006	0.0014	0.0018
	0.0810	0.0720	0.0924	0.0948
	0.7200	0.9270	0.9024	0.9048
$\hat{\alpha}_0$	0.0009	0.0011	0.0009	0.0012
	(0.0002)	(0.0002)	(0.0002)	(0.0001)
$\hat{\alpha}_1$	0.0940	0.0921	0.0891	0.0943
	(0.0012)	(0.0049)	(0.0062)	(0.0051)
$\hat{\beta}_1$	0.8996	0.8930	0.9016	0.9059
	(0.0014)	(0.0054)	(0.0065)	(0.0055)

Finally, we test the viability of our algorithm for real data. We have analyzed the Standard and Poor's 500 index for the years 1950 to 2011, which consists of 15366 daily data. We fitted a GARCH(1, 1) model, estimated the residuals, and then re-generated a GARCH(1, 1) process using both estimated parameters and residuals. Then, we computed the empirical mean of the squared error between the two processes, denoted by \hat{s} , see the last row of the table below. The results show remarkable accuracy of even such a simple model, and a good agreement in the performance of the off-line and the on-line estimator.

Table IV. Comparison of off-line and on-line estimates for S&P500 data

	Off-line	On-line
$\hat{\alpha}_0$	$7.156 \cdot 10^{-7}$ ($6.2662 \cdot 10^{-8}$)	$1.3921 \cdot 10^{-6}$ ($2.5229 \cdot 10^{-7}$)
$\hat{\alpha}_1$	0.0790 (0.0016)	0.0925 (0.0055)
$\hat{\beta}_1$	0.9156 (0.00021)	0.8981 (0.0059)
\hat{s}	$1.8632 \cdot 10^{-4}$	$2.1253 \cdot 10^{-4}$

Chapter 8

A strong approximation result for GARCH processes

The purpose of this chapter is to give a strong approximation theorem for the off-line maximum likelihood estimation for GARCH processes, following the arguments of Gerencsér [40]. Our result gives rate of convergence of $\hat{\theta}_N - \theta^*$ for basic GARCH(r, s) processes by a standard martingale approximation of this difference. The main result of this chapter is given in Section 8.1, while its proof has been presented in Section 8.2.

8.1 The main result

Recall from Chapter 4 that the conditional quasi-maximum likelihood estimation $\hat{\theta}_N$ of θ^* is defined as the solution of the equation

$$\frac{\partial}{\partial \theta} L_N(\theta, \theta^*) = L_{\theta N}(\theta, \theta^*) = 0. \quad (8.1)$$

The asymptotic cost function, (a negative log-likelihood for the gaussian case) was defined as

$$W(\theta, \theta^*) = \lim_{n \rightarrow \infty} \mathbb{E} \left[\frac{1}{2} \left(\log \bar{\sigma}_n^2(\theta) + \frac{y_n^2}{\bar{\sigma}_n^2(\theta)} \right) \right]. \quad (8.2)$$

The function $W(\theta, \theta^*)$ is smooth in $\text{int}K$ and we have

$$\frac{\partial}{\partial \theta} W_{\theta}(\theta^*, \theta^*) = W_{\theta}(\theta^*, \theta^*) = 0$$

and

$$R^* = \frac{\partial^2}{\partial \theta^2} W(\theta^*, \theta^*) = W_{\theta\theta}(\theta^*, \theta^*) > 0,$$

i.e. R^* is positive definite.

Theorem 4.1.1 of Berkes et al. [10] have shown that $N^{1/2}(\hat{\theta}_N - \theta^*)$ has the asymptotic distribution $\mathcal{N}(0, F_0^{-1}G_0F_0^{-1})$ with matrices F_0 and G_0 specified by (4.8) and (4.9). The following theorem of Berkes et al. [10] shows that the exact order of $\hat{\theta}_N - \theta^*$ is $o_P(N^{-1/2})$.

Theorem 8.1.1 (Berkes et al. [10]) *Assume that Condition 4.1.1 holds and $E|\varepsilon_0|^{2+\delta} < +\infty$ for some $\delta > 0$. Then*

$$\hat{\theta}_N - \theta^* = -(R^*)^{-1} \frac{1}{N} \sum_{n=1}^N l_{\theta n}(\theta^*, \theta^*) + r_N$$

where $r_N = o_P(N^{-1/2})$.

Now we have taken a different route: by assuming stronger conditions on the driving noise and extending the techniques for ARMA processes developed by Gerencsér [40] we provide a characterization for the error term of the off-line quasi ML estimation $\hat{\theta}_N - \theta^*$ in L_q with rate.

Theorem 8.1.2 *Let $1 - D^*(z^{-1})$ be stable, and assume further that for some positive even q*

$$\rho[E(A_0^*)^{\otimes q}] < 1 \tag{8.3}$$

is satisfied. Then

$$\hat{\theta}_N - \theta^* = -(R^*)^{-1} \frac{1}{N} \sum_{n=1}^N l_{\theta, n}(\theta^*, \theta^*) + r_N,$$

where for the error term r_N we have for $1 \leq k \leq \frac{q}{2d}$, $d > \dim \theta$,

$$E^{1/k}|r_N|^k = O(N^{-1}).$$

The main contribution of this result is that the L_k -norm of the error term r_N is shown to be $O(N^{-1})$ up to certain k -s, and the dominant term is a martingale, thus many asymptotic properties of $\hat{\theta}_N - \theta^*$ can be derived from those of martingales, such as central limit theorems and laws of iterated logarithms.

The key step of the proof, which is given in the next section, is to prove law of large numbers and uniform law of large numbers for the log-likelihood function and its derivatives with respect to θ .

8.2 Proof of Theorem 8.1.2

For the proof of the theorem we need some useful lemma.

Lemma 8.2.1 *Under the conditions of Theorem 8.1.2 we have*

$$E^{1/q'} \left| \frac{1}{N} L_N(\theta, \theta^*) - W(\theta, \theta^*) \right|^{q'} = O(N^{-1/2}),$$

with $q' = q/2$, and the same hold for the first, second and third derivatives of $W(\theta, \theta^*)$.

Proof. We first show that the process

$$u_N(\theta, \theta^*) = \frac{1}{N} L_N(\theta, \theta^*) - W(\theta, \theta^*)$$

can be reduce to the sum of a martingale difference sequence. To see this let $\nu = \nu_\theta$ be the solution of the Poisson equation

$$(I - \Pi_\theta)\nu(x) = H(\theta, x) - h(\theta)$$

with

$$h(\theta) = W(\theta, \theta^*) \quad \text{and} \quad H(\theta, \bar{X}_n^e(\theta)) = l_n(\theta, \theta^*),$$

where the state vector $\bar{X}_n^e(\theta)$ was defined by (4.16). From the application of the Poisson equation it follows that

$$u_N(\theta, \theta^*) = \frac{1}{N} \sum_{n=1}^N (H(\theta, \bar{X}_n^e(\theta)) - h(\theta)) = \frac{1}{N} \sum_{n=1}^N [\nu(\bar{X}_n^e(\theta)) - \Pi_\theta \nu(\bar{X}_n^e(\theta))].$$

Rearranging the terms in the last expression yields that $u_N(\theta, \theta^*)$ can be written as

$$\frac{1}{N} \left[\nu(\bar{X}_1^e(\theta)) + \sum_{n=1}^{N-1} [\nu(\bar{X}_{n+1}^e(\theta)) - \Pi_\theta \nu(\bar{X}_n^e(\theta))] - \Pi_\theta \nu(\bar{X}_N^e(\theta)) \right]. \quad (8.4)$$

The major observation of this expression is that the addends in the middle term form a martingale difference sequence.

Now, taking $L_{q'}$ -norm on both sides and apply the triangle inequality yields that $N E^{1/q'} |u_N(\theta, \theta^*)|^{q'}$ can be estimated from above by

$$E^{1/q'} |\nu(\bar{X}_1^e(\theta))|^{q'} + E^{1/q'} \left| \sum_{n=1}^{N-1} (\nu(\bar{X}_{n+1}^e(\theta)) - \Pi_\theta \nu(\bar{X}_n^e(\theta))) \right|^{q'} + E^{1/q'} |\Pi_\theta \nu(\bar{X}_N^e(\theta))|^{q'}$$

To see that each term in the above expression is bounded from above we need the following observations: it can be easily seen that the analysis given for the L_q -stability of the extended state vector $\bar{\psi}_n(\theta)$ in Chapter 7 is applicable for $\bar{X}_n^e(\theta)$ without changes. This implies that the statements of Lemma 7.2.1 and Lemma 7.2.3 hold for $\bar{X}_n(\theta) = \bar{X}_n^e(\theta)$ with q given by the assumption (8.3). This ensures that Condition 5.2.2 and 5.2.3 of the BMP-theory are satisfied for $\bar{X}_n(\theta) = \bar{X}_n^e(\theta)$ with q given above. On the other hand, it is straightforward to see that the function

$$H(\theta; y^2, \sigma) = \frac{1}{2} \left[\log \sigma + \frac{y^2}{\sigma} \right]$$

satisfies Condition 5.2.4 with $p = 1$. Therefore we get that Theorem 5.2.1 is applicable for $\bar{X}_n(\theta) = \bar{X}_n^e(\theta)$, and thus we have polynomial growth rate for the solution $\nu(\bar{X}_1^e(\theta))$ with $p = 1$, i.e. for all $\theta \in K$

$$|\nu(\bar{X}_1^e(\theta))| \leq C(1 + |\bar{X}_1^e(\theta)|^2)$$

with some finite constant C depending on K . Taking $L_{q'}$ -norm on both sides and applying Lemma 7.2.1 for $\bar{X}_n(\theta) = \bar{X}_n^e(\theta)$ yields the $L_{q'}$ -stability of $\nu(\bar{X}_1^e(\theta))$ with $q' = q/2$. The polynomial growth rate of the solution $\nu(\bar{X}_1^e(\theta))$ and the

first part of Condition 5.2.2 imply that for all $\theta \in K$ and $N \in \mathbb{N}$

$$|\Pi_\theta \nu(\bar{X}_N^e(\theta))| \leq C(1 + |\bar{X}_N^e(\theta)|^2)$$

with some finite constant C depending on K . By the same argument as above we get that $\Pi_\theta \nu(\bar{X}_N^e(\theta))$ is $L_{q'}$ -bounded with $q' = q/2$.

To see that the middle term in (8.4) is $L_{q'}$ -bounded first let us introduce the notation

$$m_n(\theta) = \nu(\bar{X}_{n+1}^e(\theta)) - \Pi_\theta \nu(\bar{X}_n^e(\theta)).$$

An application of the Burkholder's inequality for martingales (see Theorem 2.10 in Hall and Heyde [51]) yields the estimation

$$\mathbb{E}^{1/q'} \left| \sum_{n=1}^{N-1} m_n(\theta) \right|^{q'} \leq C \cdot \mathbb{E}^{1/q'} \left| \sum_{n=1}^{N-1} m_n^2(\theta) \right|^{q'/2}$$

with a finite constant C depending on q' . Taking the square of both sides and using the triangle inequality for the $L_{q'/2}$ norm of the right hand side we get

$$\mathbb{E}^{2/q'} \left| \sum_{n=1}^{N-1} m_n(\theta) \right|^{q'} \leq C^2 \sum_{n=1}^{N-1} \mathbb{E}^{2/q'} |m_n(\theta)|^{q'}.$$

Using the fact that $m_n(\theta)$ is a martingale difference sequence and we have polynomial growth rate for the solution $\nu(\bar{X}_{n+1}^e(\theta))$ for all n , whence the contractivity of the q -norms of the conditional expectation yields the L_q -boundedness of the process $m_n(\theta)$. This proves the first claim of the statement.

In the case of the first derivative of $W(\theta, \theta^*)$ a direct application of Lemma 7.2.1, 7.2.3 and 7.2.6 implies the statement, by the same way as above.

To prove the statement for the second derivative of $W(\theta, \theta^*)$, the first step is to extend further the state vector $\bar{X}_n^e(\theta)$ by its second derivative with respect to θ . This extensions results a linear stochastic system with block-triangular state matrix. It can be easily seen that a careful application of the same argument as above proves the statement. In the case of the third derivative of $W(\theta, \theta^*)$ a further extension of the actual state vector by the third derivative of $\bar{X}_n^e(\theta)$ results a linear stochastic system with block-triangular state matrix, as well, for

which the same argument as above is applicable. ■

Lemma 8.2.2 *Under the conditions of Theorem 8.1.2 we have*

$$\mathbb{E}^{1/q''} \left(\sup_{\theta \in K, \theta^* \in K_0} \left| \frac{1}{N} L_N(\theta, \theta^*) - W(\theta, \theta^*) \right|^{q''} \right) = O(N^{-1/2}),$$

with $q'' = q'/d$, $d > \dim \theta$ and the same hold for the first and second derivatives of $W(\theta, \theta^*)$.

Proof. The proof of this lemma follows directly from the following theorem of Gerencsér [39]:

Theorem 8.2.1 *Assume that $(x_n(\theta))$ is a stochastic process which is measurable, separable,*

$$M_r(x) = \sup_{\theta \in K, n > 0} \mathbb{E}^{1/r} |x_n(\theta)|^r < \infty,$$

for all $1 \leq r < \infty$, and the same holds for its gradient process $(x_{\theta,n}(\theta))$. Let x_n^* be the random variable defined as

$$x_n^* = \max_{\theta \in K_0} |x_n(\theta)|.$$

Then we have for all positive integer r and $d > \dim \theta$,

$$M_r(x^*) \leq C(M_{rd}(x) + M_{rd}(x_\theta))$$

where C depends only on $\dim \theta, r, d$ and K_0, K . ■

The next steps of the proof of Theorem 8.1.2 is carried out analogously to the proof of Theorem 2.1 of Gerencsér [40]. The following statements and the finish of the proof is essentially given from Gerencsér [40], appropriately modified to our setting.

Lemma 8.2.3 *For any $d > 0$ the equation (8.1) has a unique solution in D such that it is also in the sphere $\{\theta : |\theta - \theta^*| < d\}$ with probability at least $1 - O(N^{-s})$*

for any $0 < s \leq q''/2$ where the constant in the error term $O(N^{-s}) = CN^{-s}$ depends only on d and s .

Proof. We show first that the probability to have a solution outside the sphere $\{\theta : |\theta - \theta^*| < d\}$ is less than $O(N^{-s})$ with any $0 < s \leq q''/2$. Indeed, Berkes et al. [10] have shown that the equation $W_\theta(\theta, \theta^*) = 0$ has a single solution $\theta = \theta^*$ in K , thus for any $d > 0$ we have

$$d' = \inf\{|W_\theta(\theta, \theta^*)| : \theta \in D, \theta^* \in D^*, |\theta - \theta^*| \geq d\} > 0$$

since $W_\theta(\theta, \theta^*)$ is continuous in (θ, θ^*) and $K \times K_0$ is compact. Therefore if a solution of (8.1) exists outside the sphere $\{\theta : |\theta - \theta^*| < d\}$ then we have for

$$\delta L_{\theta N} = \sup_{\theta \in K, \theta^* \in K_0} \left| \frac{1}{N} L_{\theta N}(\theta, \theta^*) - W_\theta(\theta, \theta^*) \right|$$

the inequality $\delta L_{\theta N} > d'$. Due to Lemma 8.2.2 we have

$$E^{1/q''} |\delta L_{\theta N}|^{q''} = O(N^{-1/2}),$$

therefore

$$P(\delta L_{\theta N} > d') = O(N^{-s})$$

with any $0 < s \leq q''/2$ by Markov's inequality. Let us now consider the random variable

$$\delta L_{\theta\theta N} = \sup_{\theta \in K, \theta^* \in K_0} \left\| \frac{1}{N} L_{\theta\theta N}(\theta, \theta^*) - W_{\theta\theta}(\theta, \theta^*) \right\|.$$

By the same argument as above we have

$$E^{1/q''} |\delta L_{\theta\theta N}|^{q''} = O(N^{-1/2}), \quad (8.5)$$

therefore

$$P(\delta L_{\theta\theta N} > d'') = O(N^{-s})$$

for any $d'' > 0$ and hence for the event

$$A_N = \{\omega : \delta L_{\theta N} < d', \delta L_{\theta\theta N} < d''\} \quad (8.6)$$

we have for N big enough

$$P(A_N) > 1 - O(N^{-s}) \quad (8.7)$$

with any $0 < s \leq q''/2$. But on A_N the equation (8.1) has a unique solution whenever d' and d'' are sufficiently small. Indeed, the equation $W_\theta(\theta, \theta^*) = 0$ has a unique solution in K by Berkes et al. [10] and hence the existence of a unique solution of (8.1) can easily be derived from the following version of the implicit function theorem.

Lemma 8.2.4 *Let $W_\theta(\theta), \delta W_\theta(\theta)$, $\theta \in K \subset \mathbb{R}^{r+s+1}$ be \mathbb{R}^{r+s+1} -valued continuously differentiable functions, let for some $\theta^* \in K_0 \subset K$, $W_\theta(\theta^*) = 0$, and let $W_{\theta\theta}(\theta^*)$ be non-singular. Then for any $d > 0$ there exists positive numbers d', d'' such that*

$$|\delta W_\theta(\theta)| < d' \quad \text{and} \quad \|\delta W_{\theta\theta}(\theta)\| < d''$$

for all $\theta \in K_0$ implies that the equation $W_\theta(\theta) + \delta W_\theta(\theta) = 0$ has exactly one solution in a neighborhood of radius d of θ^ .*

■

Lemma 8.2.5 *We have*

$$E^{1/q''} |\hat{\theta}_N - \theta^*|^{q''} = O(N^{-1/2}).$$

Proof. Let us now consider equation (8.1) and write it as

$$0 = L_{\theta N}(\hat{\theta}_N, \theta^*) = L_{\theta N}(\theta^*, \theta^*) + \bar{V}_{\theta\theta N}(\hat{\theta}_N - \theta^*) \quad (8.8)$$

where

$$\bar{V}_{\theta\theta N} = \int_0^1 L_{\theta\theta N} \left((1-\lambda)\theta^* + \lambda\hat{\theta}_N, \theta^* \right) d\lambda.$$

First note that

$$E^{1/q''} |L_{\theta N}(\theta^*, \theta^*)|^{q''} = O(N^{1/2}), \quad (8.9)$$

by Burkholder's inequality for martingales, since $l_{\theta,n}(\theta^*, \theta^*)$ is a martingale difference sequence. Let us now investigate the integral. Define

$$\bar{W}_{\theta\theta N} = \int_0^1 W_{\theta\theta}((1-\lambda)\theta^* + \lambda\hat{\theta}_N, \theta^*)d\lambda.$$

Since the function W is smooth we have for $0 \leq \lambda \leq 1$ on the set A_N (defined in (8.6))

$$\|W_{\theta\theta}(\theta^* + \lambda(\hat{\theta}_N - \theta^*), \theta^*) - W_{\theta\theta}(\theta^*, \theta^*)\| < C|\hat{\theta}_N - \theta^*| < Cd, \quad (8.10)$$

where C is a constant depending on the system parameters. Hence if d is sufficiently small then the positive definiteness of $W_{\theta\theta}(\theta^*, \theta^*)$ implies that

$$\bar{W}_{\theta\theta N} > cI$$

with some positive c . Since on A_N

$$\left\| \frac{1}{N} \bar{V}_{\theta\theta N} - \bar{W}_{\theta\theta N} \right\| < d'',$$

it follows that if d'' is sufficiently small then

$$\lambda_{\min} \left(\frac{1}{N} \bar{V}_{\theta\theta N} \right) > c > 0$$

on A_N where $\lambda_{\min}(M)$ denotes the minimal eigenvalue of the matrix B . Hence

$$\|\bar{V}_{\theta\theta N}^{-1}\| < CN^{-1} \quad (8.11)$$

on A_N with some non-random constant C . This yields, by equation (8.8), that

$$E^{1/q''} |\chi_{A_N}(\hat{\theta}_N - \theta^*)|^{q''} = O(N^{-1/2}). \quad (8.12)$$

Combining this inequality with inequality (8.7), and using the fact that $|\hat{\theta}_N - \theta^*|$ is bounded we have

$$E^{1/q''} |\chi_{A_N^c}(\hat{\theta}_N - \theta^*)|^{q''} = O(N^{-s}) \quad (8.13)$$

with any $0 < s \leq q''/2$. Adding this inequality to (8.12) we get the lemma. \blacksquare

Now we can complete the proof of Theorem 8.1.2 as follows. According to Lemma 8.2.5 the $L_{q''}$ -norm of the inequality (8.10) can be improved by $O(N^{-1/2})$ on the right hand side. Thus we get after integration with respect to λ that

$$E^{1/q''} \left\| \bar{W}_{\theta\theta N} - W_{\theta\theta}(\theta^*, \theta^*) \right\|^{q''} = O(N^{-1/2}).$$

On the other hand the inequality $E^{1/q''} |\delta L_{\theta\theta N}|^{q''} = O(N^{-1/2})$ implies that

$$E^{1/q''} \left\| \frac{1}{N} \bar{V}_{\theta\theta N} - \bar{W}_{\theta\theta N} \right\|^{q''} = O(N^{-1/2}).$$

Hence we finally get

$$E^{1/q''} \left\| \frac{1}{N} \bar{V}_{\theta\theta N} - W_{\theta\theta}(\theta^*, \theta^*) \right\|^{q''} = O(N^{-1/2}).$$

Considering that on A_N (8.11) is satisfied and the fact that $W_{\theta\theta}(\theta^*, \theta^*) > 0$ we have

$$E^{1/q''} \left(\chi_{A_N} \left\| \bar{V}_{\theta\theta N}^{-1} - \frac{1}{N} W_{\theta\theta}^{-1}(\theta^*, \theta^*) \right\| \right)^{q''} = O(N^{-3/2}). \quad (8.14)$$

Now we get our final estimate for $\hat{\theta}_N - \theta^*$ by considering (8.8) on the set A_N . We have

$$\chi_{A_N}(\hat{\theta}_N - \theta^*) = -\chi_{A_N} \bar{V}_{\theta\theta N}^{-1} L_{\theta N}(\theta^*, \theta^*).$$

Taking into account estimation (8.14) we get that

$$\chi_{A_N}(\hat{\theta}_N - \theta^*) = -\chi_{A_N} \left(\frac{1}{N} W_{\theta\theta}^{-1}(\theta^*, \theta^*) + v_N \right) L_{\theta N}(\theta^*, \theta^*),$$

where the error term v_N is such that $E^{1/q''} |v_N|^{q''} = O(N^{-3/2})$, and (8.9) implies that

$$\chi_{A_N}(\hat{\theta}_N - \theta^*) = -\chi_{A_N} \frac{1}{N} W_{\theta\theta}^{-1}(\theta^*, \theta^*) L_{\theta N}(\theta^*, \theta^*) + w_N$$

with $E^{1/q''} |w_N|^{q''} = O(N^{-1})$. Considering (8.7) and (8.9) we have that the $L_{q''}$ -norm of

$$(1 - \chi_{A_N}) \frac{1}{N} W_{\theta\theta}^{-1}(\theta^*, \theta^*) L_{\theta N}(\theta^*, \theta^*)$$

is $O(N^{-1/2})$, which implies that

$$\chi_{A_N}(\hat{\theta}_N - \theta^*) = W_{\theta\theta}^{-1}(\theta^*, \theta^*) \frac{1}{N} L_{\theta N}(\theta^*, \theta^*) + z_N,$$

with $E^{1/q''} |z_N|^{q''} = O(N^{-1})$. Combining this with (8.13) and using the definition of $W_{\theta\theta}^{-1}(\theta^*, \theta^*)$ and $L_{\theta N}(\theta^*, \theta^*)$ we get the proposition of the theorem.

Chapter 9

Conclusions and further developments

In this thesis, a new approach to the recursive estimation of GARCH processes has been proposed which is a natural adaptation of the off-line quasi-maximum likelihood method. It can also be considered as an extension of the recursive prediction error identification method widely studied in the theory of linear stochastic systems. Surprisingly, the theoretical analysis of this method requires a deep mathematical technology developed in Benveniste et al. [7]. Convergence in almost sure sense and in L_q has been established under reasonable and verifiable technical conditions, with the exception of (7.4). The best alternative to our method is the two pass recursive estimation method due to Aknouche and Guerbyenne [1]. However, it has a limited range of applicability, as well, due to the positive real condition (7.7). More importantly, the technical conditions under which the results are valid are not fully specified. In particular, the controversial "boundedness condition", ensuring that the estimators of β_i stay inside a pre-specified domain, is not addressed. The asymptotic covariance matrices of the estimators are believed to be the same as that for the off-line estimator, however the existing arguments seem to be incomplete.

Our method exploits the relatively simple structure of the dynamics of the model, allowing its inversion. This observation implies further applicability of our method. A direct extension of our results to several generalizations of GARCH model, such as the EGARCH, TGARCH and PGARCH models dis-

cussed in Section 2.2, is partially possible, since many of them can be formulated in the form of the linear stochastic system (3.6). In more complex stochastic volatility models we may be forced to compute the likelihood of a specific observation. In such circumstance smart Monte Carlo techniques may be useful. It is an interesting question whether to convert this technique into a recursive method is possible. A further application would be the class of general bilinear models, introduced by Granger and Andersen [49]. It is easy to see that the so-called sub-diagonal bilinear model defined by

$$y_t = \sum_{j=1}^p a_j y_{t-j} + e_t + \sum_{j=1}^r b_j e_{t-j} + \sum_{k=1}^R \sum_{j=0}^P b_{kj} y_{t-j-k} e_{t-k} \quad (9.1)$$

where (e_t) is a sequence of i.i.d. random variables with zero mean and variance σ^2 , can be written in the form of model (3.6) with (A_n, u_n) being an i.i.d. sequence of random variables satisfying Condition 3.3.1, see Pham [77]. Note, that for even $q \geq 2$ the condition $\rho[E(A_1^{\otimes q})] < 1$ shows up in a necessary and sufficient condition for the existence of q -th order moments, see Terdik [88], Theorem 45.

There seems to be one more model class of research that has important promise for application. These are high dimension multivariate models, especially the so-called factor-models. In this thesis, we have focused on modelling the volatility of a single asset. However, for most financial applications, there are thousands of assets. During the last 20 years multivariate models of financial assets have received considerable attention in the literature. Many researchers are already developing several different multivariate GARCH models for modelling the conditional covariance matrix. Unfortunately, the parametrization of the conditional covariance matrix is rather complicate because of the high number of the parameters and the requirements for the positive definiteness of the covariance matrix. One of the most practically useful model of these specifications, motivated by economic theory, is the so-called Factor-GARCH model developed by Engle et al. [33]. In this model it is assumed that the conditional covariance matrix H_n is generated by K factors $f_{k,n}$ ($K < N$ with N denoting

the number of assets) as

$$H_n = \Omega + \sum_{k=1}^K w_k w_k^T f_{k,n}, \quad n \in \mathbb{Z},$$

where Ω is an $N \times N$ positive definite matrix, w_k , $k = 1, \dots, K$ are the factor weights, and the factors follow univariate GARCH structure, respectively. The model has the advantage that it reduces the dimension of the model when the number of the factors relative to N is small.

A second effort of this research was to develop strong approximation result, which provide representation of the error process of the off-line maximum likelihood estimator in a very convenient form. A key element in the analysis was a uniform strong law of large numbers for the log-likelihood function and its derivatives.

It is also an important question that within the examined time period the financial time series have been change in its stochastic structure. If yes then where these changes are happened. A further development of our research would be the detection of the changes of the parameters in the GARCH model using the theory of the so-called fixed gain estimation, where the estimation procedure is modified by using exponential forgetting in the off-line case.

Appendix A

Basic properties of Kronecker products

Here we summarize the definition and basic properties of Kronecker products. Let $A \in \mathbb{R}^{n_1 \times n_2}$ and $B \in \mathbb{R}^{m_1 \times m_2}$ be matrices of arbitrary dimensions. Then their Kronecker product is an $(n_1 m_1) \times (n_2 m_2)$ matrix defined by

$$A \otimes B = \begin{pmatrix} a_{11}B & \dots & a_{1,n_2}B \\ \vdots & & \vdots \\ a_{n_1,1}B & \dots & a_{n_1,n_2}B \end{pmatrix}.$$

The Kronecker power of a matrix is defined by

$$A^{\otimes k} = \underbrace{A \otimes \dots \otimes A}_{k \text{ times}}.$$

There is no need for parentheses in this expression since the Kronecker product is associative, as stated below. We will also need the notation $\text{vec}M$ denoting the vector that is obtained by the concatenating the columns of the matrix M . The following lemma summarizes some elementary properties of the Kronecker product.

Lemma A.0.6 *We have*

(i) $A \otimes (B \otimes C) = (A \otimes B) \otimes C$

(ii) $(A \otimes B)^T = A^T \otimes B^T$

(iii) *for matrices of compatible dimensions we have*

$$(A_1 \otimes B_1) \cdot (A_2 \otimes B_2) = (A_1 \cdot A_2) \otimes (B_1 \cdot B_2)$$

(iv) *for square matrices we have $\text{tr}(A \otimes B) = \text{tr}A \cdot \text{tr}B$*

(v) $\|A^{\otimes k}\| = \|A\|^k$

(vi) *we have for matrices of compatible dimensions*

$$\text{vec}(AX) = (I \otimes A) \cdot \text{vec}X \quad \text{and} \quad \text{vec}(XC^T) = (C \otimes I) \cdot \text{vec}X$$

(vii) *for square matrices A we have*

$$\text{tr}A = \text{vec}I^T \cdot \text{vec}A,$$

where I is a unit matrix of the same dimension.

For the details of the proof we refer the reader to [6, 13, 79].

Appendix B

Proof of Lemma 7.2.4 and 7.2.5

Proof of lemma 7.2.4: Applying Lemma 7.2.5 in (7.19) yields that the first term of (7.19) can be estimated from above by

$$\bar{C}\bar{\rho}^n|\theta_1 - \theta_2|$$

with $\bar{\rho} = \max(\rho, \rho') \in (0, 1)$ and some finite constant \bar{C} . To estimate the second term of (7.19) let us recall that

$$U_n(\theta_i) = \sum_{k=0}^{n-1} P_n(\theta_i) \dots P_{n-k+1}(\theta_i) w_{n-k}(\theta_i), \quad i = 1, 2.$$

Applying now the triangle inequality yields $E^{1/l}|U_n(\theta_1) - U_n(\theta_2)|^l$ can be estimated from above by

$$\begin{aligned} & \sum_{k=0}^{n-1} E^{1/l} \left\| \prod_{i=0}^{k-1} P_{n-i}(\theta_1) - \prod_{i=0}^{k-1} P_{n-i}(\theta_2) \right\|^l \cdot E^{1/l} \|w_{n-k}(\theta_1)\|^l + \\ & + \sum_{k=0}^{n-1} E^{1/l} \left\| \prod_{i=0}^{k-1} P_{n-i}(\theta_2) \right\|^l \cdot E^{1/l} \|w_{n-k}(\theta_1) - w_{n-k}(\theta_2)\|^l. \end{aligned} \tag{B.1}$$

To estimate the elements of the first term of (B.1) we apply Lemma 7.2.4. This yields that there exist positive constants $C_1, K_1 < +\infty$ and $0 < \rho < 1$ such that

the first part of (B.1) can be estimated from above by

$$\sum_{k=0}^{n-1} C_1 K_1 \rho^k |\theta_1 - \theta_2|. \quad (\text{B.2})$$

To estimate the elements of the second term of (B.1) we use Corollary 7.1.1 and the fact that $w_k(\theta)$ is smooth in θ . Thus we get that there exist positive constants $C_2, K_2 < +\infty$ and $0 < \rho' < 1$ such that the second part of (B.1) can be estimated from above by

$$\sum_{k=0}^{n-1} C_2 K_2 (\rho')^k |\theta_1 - \theta_2|. \quad (\text{B.3})$$

Taking into account (B.1)-(B.3) yields

$$E^{1/l} |U_n(\theta_1) - U_n(\theta_2)|^l \leq \bar{C} |\theta_1 - \theta_2|$$

with some finite constant \bar{C} , which proves the statement.

Proof of Lemma 7.2.5: Let us observe first that if (A_i) and (B_i) are sequences of square matrices then

$$\prod_{i=1}^n A_i - \prod_{i=1}^n B_i = \sum_{k=1}^n A_n \dots A_{k+1} (A_k - B_k) B_{k-1} \dots B_1.$$

Using the above fact with $A_i = P_i(\theta_1)$ and $B_i = P_i(\theta_2)$ and the triangle inequality yields that the elements of the first term of (B.1) can be estimated from above by

$$\sum_{k=1}^n E^{1/l} \|P_n(\theta_1) \dots P_{k+1}(\theta_1)\|^l E^{1/l} \|P_k(\theta_1) - P_k(\theta_2)\|^l E^{1/l} \|P_{k-1}(\theta_2) \dots P_1(\theta_2)\|^l. \quad (\text{B.4})$$

By Corollary 7.1.1 it follows that there exist constants C_1, C_2 and $\rho, \rho' \in (0, 1)$

such that (B.4) can be estimated from above by

$$\sum_{k=1}^n C_1 \rho^{n-k} \mathbb{E}^{1/l} \|P_k(\theta_1) - P_k(\theta_2)\|^l C_2(\rho')^k.$$

Since $P_k(\theta)$ is smooth in θ it follows that there exists a finite constant K such that

$$\|P_k(\theta_1) - P_k(\theta_2)\| \leq K|\theta_1 - \theta_2|.$$

Thus we get that, for a sufficiently large n , (B.4) can be bounded from above by

$$\bar{C} \bar{\rho}^n |\theta_1 - \theta_2|$$

with $\bar{\rho} = \max(\rho, \rho') \in (0, 1)$ and some finite constant \bar{C} , from which the claim follows.

Bibliography

- [1] A. Aknouche and H. Guerbyenne, Recursive estimation of GARCH models, *Communication in Statistics-Simulation and Computation* 35 (2006) 925-938.
- [2] L. Arnold, A formula connecting sample and moment stability of linear stochastic systems, *SIAM J. Appl. Math.* 44(4) (1984) 793-802.
- [3] L. Arnold, W. Kliemann and E. Oeljeklaus, Lyapunov exponents of linear stochastic systems, In: *Lyapunov Exponents, Lecture Notes in Mathematics* 1186, Springer-Verlag, 1984 (Eds.: L. Arnold and V. Wihstutz)
- [4] N.E. Barabanov, On the Lyapunov exponent of discrete inclusions. I–III, *Automation and Remote Control*, 49 (1988) 152-157, 283-287, 558-565.
- [5] N.E. Barabanov, Lyapunov exponent and joint spectral radius: some known and new results, In: *Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference 2005 Seville, Spain, December 12-15, 2005*, TuA08.1, pp. 2332-2337
- [6] R. Bellman, *Introduction to matrix analysis*, McGraw-Hill, New York, 1960.
- [7] A. Benveniste, M. Métivier and P. Priouret, *Adaptive Algorithms and Stochastic Approximations*, Springer Verlag, Berlin, 1990.
- [8] I. Berkes and L. Horváth, The rate of consistency of the quasi-maximum likelihood estimator, *Statistics and Probability Letters* 61 (2003) 133-143.

- [9] I. Berkes and L. Horváth, The efficiency of the estimators of the parameters in GARCH processes, *The Annals of Statistics* 32(2) (2004) 633-655.
- [10] I. Berkes, L. Horváth, P.S. Kokoszka, GARCH processes: structure and estimation, *Bernoulli*, 9 (2003) 201-217.
- [11] F. Black, Studies of stock price volatility changes, In *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economic Statistics Section* (1976) 177-181.
- [12] F. Black and M. Scholes, The valuation of option contracts and a test of market efficiency, *Journal of Finance*, 27(2) (1972) 399-417.
- [13] V.D. Blondel and Y. Nesterov. Computationally efficient approximations of the joint spectral radius, *SIAM Journal on Matrix Analysis and Applications*, 27(1) (2005) 256-272
- [14] T. Bollerslev, Generalized autoregressive conditional heteroscedasticity, *Journal of Econometrics*, 31 (1986) 307-327.
- [15] T. Bollerslev, R.F. Engle and D.B. Nelson, ARCH models. In R.F. Engle and D.L. McFadden, editors, *Handbook of Econometrics Vol.4*, Elsevier Science B.V., Amsterdam, 1994.
- [16] P. Bougerol, N. Picard. Stationarity of GARCH processes and of some nonnegative time series, *Journal of Econometrics*, 52 (1992) 115-127.
- [17] P. Bougerol, N. Picard. Strict stationarity of generalized autoregressive processes, *The Annals of Probability*, 20(4) (1992) 1714-1730.
- [18] G.E.P. Box and G.M. Jenkins, *Time Series Analysis, Forecasting and Control*, Wiley, 1976.
- [19] A. Brandt, The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients, *Adv. Appl. Prob.* 18 (1986) 211-220.
- [20] M. Carrasco, X. Chen, Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* 18(1) (2002) 17-39.

- [21] K.S. Chan and H. Tong, On the use of deterministic Lyapunov function for the ergodicity of stochastic difference equations, *Adv. Appl. Prob.* 17 (1985) 666-678.
- [22] Y.S. Chow, H. Teicher, Independence, Interchangeability, Martingales, Springer Verlag, 1978.
- [23] R. Cont, Empirical properties of asset returns: Stylized facts and statistical issues, *Quantitative Finance* 1 (2001) 1-14.
- [24] R. Cont, Volatility clustering in financial markets: empirical facts and agent-based models, In: Long memory in economics, 2005 (eds.: A. Kirman and G. Teyssiere).
- [25] R. Dahlhaus and S. Subba Rao, A recursive online algorithm for the estimation of time-varying ARCH parameters, *Bernoulli* 13(2) (2007) 389-422.
- [26] J. Danielsson and J.-F. Richard, Accelerated Gaussian Importance Sampler with Application to Dynamic Latent Variable Models, *Journal of Applied Econometrics* 8 (1993) 153-173.
- [27] W. Dayawansa, C.F. Martin, A converse Lyapunov theorem for a class of dynamical systems which undergo switching, *IEEE Trans. Automat. Contr.* 44 (1999) 751-760.
- [28] B. Delyon, Stochastic approximation with decreasing gain: Convergence and asymptotic theory, technical report, Université de Rennes, 2000. s
- [29] Z. Ding, C.W.J. Granger and R.F. Engle, A long memory property of stock market returns and a new model, *Journal of Empirical Finance* 1 (1993) 83-106.
- [30] P. Doukhan. Mixing: Properties and Examples. Springer Verlag, New York, 1994.
- [31] R.F. Engle, Autoregressive conditional heteroscedasticity with estimates of the variance of the United Kingdom inflation, *Econometrica*, 50 (1982) 987-1008.

- [32] R.F. Engle, Risk and volatility: econometrics models and financial practice, *The American Economic Review* 94(3) (2004) 405-420.
- [33] R.F. Engle, V. Ng and M. Rothschild, Asset pricing with a Factor ARCH covariance structure: empirical estimates for Treasury Bills, *Journal of Econometrics* 45 (1990) 213-238.
- [34] Y. Fang and K.A. Loparo, On the relationship between the sample path and moment Lyapunov exponents for jump linear systems, *IEEE Trans. Automat. Contr.* 47(9) (2002) 1556-1560.
- [35] P.D. Feigin and R.L. Tweedie, Random coefficient autoregressive processes: a Markov chain analysis of stationarity and finiteness of moments, *Journal of Time Series Analysis*, 6(1) (1985) 1-14.
- [36] P.H. Franses and D. van Dijk, *Non-linear time series models in empirical finance*, Cambridge University Press, Cambridge, 2000.
- [37] H. Furstenberg and H. Kesten, Products of random matrices, *Ann. Math. Statist.* 31 (1960) 457-469.
- [38] H. Furstenberg and Y. Kifer, Random matrix products and measures on projective spaces, *Israel J. Math.* 46(1-2) (1983) 12-32.
- [39] L. Gerencsér. On a class of mixing processes. *Stochastics* 26 (1989) 165-191.
- [40] L. Gerencsér. On the Martingale approximation of the estimation error of ARMA parameters. *Systems & Control Letters* 15 (1990) 417-423.
- [41] L. Gerencsér. Rate of convergence of recursive estimators. *SIAM Journal Control and Optimization* 30(5) (1992) 1200-1227.
- [42] L. Gerencsér. A representation theorem for the error of recursive estimators, *SIAM Journal on Control and Optimization* 44(6) (2006) 2123-2188.
- [43] L. Gerencsér and Z. Mátyás (2007). Almost Sure and L_q Convergence of the Re-Initialized BMP scheme, In: *Proc. of the 46th Conference on Decision and Control CDC'07*, New Orleans, LA, USA, Dec. 12-14, 2007, WeB11.4, pp. 969-974.

- [44] L. Gerencsér, Gy. Michaletzky and Zs. Orlovits, Stability of block-triangular stationary random matrices, *Systems and Control Letters* 57 (2008) 620-625.
- [45] L. Gerencsér and Zs. Orlovits, L_q -stability of products of block-triangular stationary random matrices, *Acta Scientiarum Mathematicarum* (Szeged) 74 (2008) 927-944.
- [46] L. Gerencsér and Zs. Orlovits, Real time estimation of stochastic volatility processes, *Annals of Operations Research, Special Volume: Stochastic Modeling and Optimization. Dedicated to the 80th Birthday of Professor András Prékopa*. Accepted for publication.
- [47] L. Gerencsér, Zs. Orlovits and B. Torma, Recursive estimation of GARCH processes, In *Proceedings of the 19-th International Symposium on Mathematical Theory of Networks and Systems (MTNS 2010)* (eds: L. Gerencsér, Gy. Michaletzky and A. Edelmayer), Budapest, 5-9 July 2010, CD ROM, ISBN 978-963-311-370-7, pp. 2415- 2422.
- [48] L.R. Glosten, R. Jagannathan and D.E. Runkle, On the relation between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance* 48(5) (1993) 1779-1801.
- [49] C.W.J. Granger, A. Andersen, *An introduction to Bilinear Time Series Model*, Vanderhoeck and Ruprecht, Gottingen, 1978.
- [50] D. Guegan and J. Diebolt, Probabilistic properties of the β -ARCH model, *Statistica Sinica* 4 (1994) 71-87.
- [51] D. Hall and C.C. Heyde, *Martingale Limit Theory and Application*, Academic Press, New York, 1980.
- [52] E. Hannan, The asymptotic theory of linear time-series models, *Journal of Applied Probability* 10 (1973) 130-145.
- [53] R.Z. Hasminskii, *Stochastic stability of differential equations*, Sijthoff and Noordhoff, Alphen aan den Rijn, The Netherlands, 1980.

- [54] C. He and T. Terasvirta, Properties of moments of a family of GARCH processes, *Journal of Econometrics* 92 (1999) 173-192.
- [55] M. Karanasos, The second moment and the autocovariance function of the squared errors of the GARCH model, *Journal of Econometrics* 90 (1999) 63-76.
- [56] H.A. Karlsen, Existence of moments in a stationary stochastic difference equation, *Adv. Appl. Prob.* 22 (1990) 129-146.
- [57] H. Kesten and F. Spitzer, Convergence in distribution of products of random matrices, *Zeitschrift für Wahrscheinlichkeit und verwandete Gebiete*, Vol. 67 (1984) pp. 363 -386.
- [58] J.L. Kierkegaard, J.N. Nielsen, L. Jensen and H. Madsen, Estimating GARCH models using recursive methods, <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.40.703>, (2000).
- [59] H.J. Kushner and D.S. Clark, *Stochastic Approximation Methods for Constrained and Unconstrained Optimization*, Springer-Verlag, Berlin, New York, 1978.
- [60] S. Lee and B. Hansen. Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory* 10 (1994) 29-52.
- [61] S. Ling, On the probabilistic properties of a double threshold ARMA conditional heteroscedasticity model, *Journal of Applied Probability* 36 (1999) 688-705.
- [62] S. Ling, M. McAleer, Necessary and sufficient moment conditions for the GARCH(p,q) and asymmetric power GARCH(p,q) models, *Econometric Theory* 18 (2002) 722-729.
- [63] L. Ljung, Analysis of recursive stochastic algorithms. *IEEE Trans. on Automatic Control* 22 (1977) 551-575.
- [64] L. Ljung and T. Söderström, *Theory and practice of recursive identification*, MIT Press, 1987.

- [65] R. Lumsdaine, Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1, 1) and covariance stationary GARCH(1, 1) models. *Econometric Theory* 64 (1996) 575-596.
- [66] H.M. Markowitz, Portfolio Selection, *Journal of Finance* 7(1) (1952) 77-91.
- [67] R.C. Merton Theory of rational options pricing, *Bell Journal of Economics and Management Science* 4(1) (1973) 141-183.
- [68] S.P. Meyn, R.L. Tweedie. *Markov Chains and Stochastic Stability*. Springer Verlag, London, 1993.
- [69] Gy. Michaletzky, L. Gerencsér, BIBO stability of linear switching systems, *IEEE Trans. Automat. Contr.* 47(11) (2002) 1895-1898.
- [70] T. Mikosch, D. Straumann, Stable limit of martingale transforms with application to the estimation of GARCH parameters, *Annals of Statistics*, 34(1) (2006) 493-522.
- [71] T. Mikosch, D. Straumann, Quasi-maximum likelihood estimation in heteroskedastic time series: a stochastic recurrence equation approach,. *Annals of Statistics* 34(5) (2006) 2449-2495.
- [72] A. Milhøj, The moment structure of ARCH processes, *Scandinavian Journal of Statistics* 12 (1985) 281-292.
- [73] D.B. Nelson, Conditional heteroscedasticity in asset returns: a new approach, *Econometrica* 59(2) (1991) 347-370.
- [74] D.F. Nicholls, B.G. Quinn, *Random coefficient autoregressive models: An introduction*, Springer-Verlag Lecture Notes in Statistics Volume 11: New-York, 1982.
- [75] V.I. Oseledec, A multiplicative ergodic theorem. Lyapunov characteristic numbers for dynamical systems, *Trans. Moscow Math. Soc.* 19 (1968) 197-231.
- [76] A. Pagan, The econometrics of financial markets, *Journal of Empirical Finance* 3 (1986) 15-102.

- [77] D.T. Pham, Bilinear Markovian representation and bilinear models, *Stochastic Processes and their Applications* 20 (1985) 295-306.
- [78] D.T. Pham, The mixing property of bilinear and generalised random coefficient autoregressive models, *Stochastic Processes and their Applications* 23 (1986) 291-300.
- [79] V.V. Prasolov, *Problems and Theorems in Linear Algebra*, Translation of Mathematical Monographs Vol.134, American Mathematical Society, Providence, RI, 1994.
- [80] M. S. Ragunathan. A proof of Oseledec's multiplicative ergodic theorem, *Israel Journal of Mathematics*, 42 (1979) 356-362.
- [81] G.C. Rota and G. Strang, A note on the joint spectral radius, *Indag. Math.* 22 (1960) 379-381.
- [82] W. Sharp, Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk, *Journal of Finance* 19(3) (1964) 425-442.
- [83] D. Stefanescu, New bounds for positive roots of polynomials, *J. Univ. Comp. Sc.* 11 (2005) 2125-2131.
- [84] D. Straumann, *Estimation in Conditionally Heteroscedastic Time Series Models*, volume 181 of *Lecture Notes in Statistics*. Springer-Verlag, Berlin, 2005.
- [85] S. Subba Rao, On some nonstationary, nonlinear random processes and their stationary approximations, *Advances in Applied Probability* 38 (2006) 1155-1172.
- [86] S. Taylor, *Modelling financial time series*, New York, Wiley, 1986.
- [87] T. Terasvirta, An introduction to univariate GARCH models, In *Handbook of Financial Time Series*, New York, Springer, (eds.: T.G. Andersen, R.A. Davis, J-P. Kreiss and T. Mikosch) 2008.

- [88] Gy. Terdik, Bilinear Stochastic Models and Related Problems of Nonlinear Time Series Analysis, volume 142 of Lecture Notes in Statistics, Springer Verlag, New York (1999).
- [89] H. Tong Non-linear Time Series: A Dynamical System Approach New York, Oxford University Press, 1990.
- [90] J.N. Tsitsiklis, The stability of the products of a finite set of matrices, in Open Problems in Communication and Computation, Springer-Verlag, New York, (eds.: T.M.Cover and B.Gopinath) 1987, 161-163.
- [91] R.L. Tweedie, Criteria for rates of convergence of Markov chains, with applications to queueing and storage theory, in Probability, Statistics and Analysis (eds.: J.F.C. Kingman and G.E.H. Reuter), Cambridge University Press, Cambridge, 1983.
- [92] R.L. Tweedie, Invariant measure for Markov chains with no irreducibility assumptions Journal of Applied Probability 25A (1988) 275-285.
- [93] W. Vervaat, On a stochastic difference equation and a representation of nonnegative infinitely divisible random variables, Adv. Appl. Prob. 11 (1979) 750-783.
- [94] A.A. Weiss, Asymptotic Theory for ARCH Models: Estimation and Testing, Econometric Theory 2 (1986) 107-131.
- [95] P.A. Zadrozny, Necessary and sufficient restrictions for existence of a unique fourth moment of a univariate GARCH(p, q) process, Advances in Econometrics 20(1) (2006) 365-379.
- [96] M. Zakoian, Threshold heteroscedastic models, Journal of Economic Dynamics and Control 18 (1994) 931-955.

Summary

The main objective of this thesis is to introduce and analyse an appropriate recursive estimation method for the parameters of the generalized autoregressive conditional heteroscedastic (GARCH) model, and to present a strong approximation result for the error term of the off-line maximum likelihood estimator.

The key observation of the analysis of GARCH model is that its specific structure allows us to define a special linear state-space model which ensures Markovian structure. The rigorous analysis of these linear dynamics helps us to understand GARCH models in detail and to find conditions for the higher order properties of these models. A new result connected to the higher order properties of general linear stochastic systems is presented.

The key problem of the statistical analysis of GARCH model is the estimation of the parameters. In this thesis we construct a recursive or on-line algorithm for estimating the parameters of the GARCH process in real time. The construction and the convergence property of the proposed algorithm is based on the theory of stochastic approximation with Markovian dynamics presented in the book of Benveniste, Métivier and Priouret (1990), shortly called as BMP-scheme. In order to reformulate the problem of estimating the GARCH parameters into a form that fits the framework of the BMP theory we define an extended state-space equation for the inverse system, and its derivative with respect to the parameters. This extension results a linear stochastic system with block-triangular state matrix. For the applicability of the BMP-theory we have developed two useful technical tools: we have provided a simple method for the computation of the top-Lyapunov exponent of block-triangular matrices and of the L_q -stability of block-triangular random matrix products. Applying the BMP-theory and a suitable resetting mechanism we prove the almost sure convergence of the proposed algorithm, using the results of Gerencsér and Mátyás (2007), and also L_q convergence up to certain q -s. The convergence of our algorithm is demonstrated by experimental result both for simulated and real data.

Finally we worked out a strong approximation result for the error process of the off-line quasi-maximum likelihood estimator.

Összefoglaló

A disszertáció fő célja az ún. GARCH (általánosított autoregresszív feltételese heteroszkedasztikus) modell paramétereinek valós idejű becslése, a javasolt rekurzív becslési algoritmus statisztikai elemzése, illetve egy erős approximációs eredmény kidolgozása az off-line maximum likelihood becslés hibájára.

A GARCH folyamatok elemzésének legfontosabb észrevétele az, hogy a modell speciális struktúrája lehetővé teszi a folyamat lineáris állapottér felírását, mely Markov struktúrát biztosít a folyamat számára. Ezen lineáris dinamika vizsgálata segíti a GARCH folyamatok részletes elemzését és a modell magasabbrendű momentumainak létezésére vonatkozó feltételek vizsgálatát. A lineáris sztochasztikus rendszer magasabbrendű momentumaival kapcsolatban új eredményt fogalmazunk meg.

A GARCH folyamatok statisztikai elemzésének legfontosabb feladata a modell paramétereinek becslése. A disszertációban rekurzív (on-line) algoritmust konstruálunk a GARCH folyamat paramétereinek valós idejű becslésére. Az algoritmus konstrukciója és konvergenciájának elemzése a Benveniste, Metivier és Priouret (1990), a továbbiakban BMP, könyvében kidolgozott, Markov dinamikán alapuló sztochasztikus approximáció elméletén alapul. Annak érdekében, hogy a GARCH folyamatok paraméterbecslési problémáját olyan formában tudjuk megfogalmazni, mely illeszkedik a BMP elméletbe, egy ún. kibővített állapotot egyenletet definiálunk az invertált folyamatra és annak deriváltjára. Ez a bővítés olyan lineáris sztochasztikus rendszert eredményez, melynek átmenetmátrixa blokk-háromszög struktúrájú. A BMP elmélet alkalmazásához két hasznos technikai eszközt dolgozunk ki: egyszerű eljárást adunk blokk-háromszög mátrixok top-Lyapunov exponensének kiszámítására, illetve véletlen blokk-háromszög mátrixok szorzatának L_q -stabilitására vonatkozóan. A BMP elmélet eszköztárának és Gerencsér és Mátyás (2007) ún. újraindítási módszerének felhasználásával a javasolt rekurzív algoritmus 1 valószínűségű konvergenciáját igazoljuk. A kidolgozott algoritmus konvergenciáját szimulált GARCH adatokon és valós adatokon is teszteljük.

Végül erős approximációs tételt dolgozunk ki az off-line maximum-likelihood becslés hibájára vonatkozóan.